

Sentimen Analisis Pelanggan Shopee di Twitter menggunakan Algoritma Naive Bayes

Aditya Hastami Ruger¹, M Suyanto², Mei P Kurniawan³

^{1,2,3}Teknologi Informatika, Universitas AMIKOM Yogyakarta

¹Aditya.ruger@students.amikom.ac.id, ²yanto@amikom.ac.id, ³meikurniawan@amikom.ac.id

Abstract - Sentiment analysis is a field of study that analyzes a person's opinions, evaluates, judges, behaviors, and emotions such as products, services, events, and topics. Sentiment analysis in the business world is usually used to analyze the needs of the community and market needs, which are expected to be able to develop marketing strategies that can increase their company's income. This study takes the data from the Shopee group on Twitter. Shopee is one of the market places that are often used by the people of Indonesia and people in other countries. Shopee's market place that sells services or sells goods for daily needs such as electronics, cutlery, credit, as well as airline tickets, train tickets, and many more services sold by Shopee. The purpose of this research is to find out how many reviews of positive comments and negative comments on the Shopee group on Twitter social media, the data collection uses Rstudio. The Rstudio application can be run on Windows, Linux, or Apple operating systems. As well as for the calculation process using the Naive Bayes method which is included in the sentiment analysis where data collection is taken from the Shopee group. In calculating the level of accuracy using the confusion matrix. The results of this study are that it can be seen that the number of positive and negative comments is equal, namely 150:150, because the data taken is 300 data and the level of accuracy is 97%.

Intisari - Sentimen analisis merupakan suatu bidang studi yang menganalisis tentang opini-opini seseorang, mengevaluasi, menilai, perilaku, dan emosi seperti produk-produk, layanan, kejadian, serta topik-topik. Sentimen analisis dalam dunia bisnis biasanya dipakai untuk menganalisis suatu kebutuhan masyarakat dan kebutuhan pasar, yang diharapkan mampu menyusun strategi pemasaran yang dapat meningkatkan pendapatan perusahaan mereka. Penelitian ini mengambil data-datanya dari grup Shopee yang ada di Twitter. Shopee merupakan salah satu market place yang sering digunakan oleh masyarakat Indonesia maupun masyarakat dinegara lain. Market place Shopee yang menjual jasa maupun menjual barang-barang untuk kebutuhan sehari-hari seperti elektronik, alat-alat makan, pulsa, maupun tiket pesawat, tiket kereta, dan masih banyak lagi jasa yang dijual oleh Shopee. Tujuan dari dilakukan penelitian ini yaitu untuk mengetahui berapa jumlah review dari komentar positif maupun komentar negatif pada grup Shopee yang ada di media sosial Twitter, pengambilan datanya menggunakan Rstudio, Rstudio merupakan suatu aplikasi yang dipakai untuk menulis program-program dengan menggunakan bahasa R. Aplikasi Rstudio dapat dijalankan pada operasi sistem Windows, Linux, maupun Apple. Serta untuk proses menghitung menggunakan metode Naive Bayes yang dimasukkan kedalam sentimen analisis yang pengambilan data di ambil dari grup Shopee. Dalam penghitungan tingkat keakurasi dengan menggunakan confusion matriks. Hasil dari penelitian ini yaitu dapat diketahui bahwa jumlah komentar positif maupun negatif setara yaitu 150:150, karena data yang diambil adalah 300 data dan tingkat keakurasiannya adalah sebesar 97%.

Kata kunci - Sentimen Analisis, Naive Bayes, Shopee, Rstudio, Confusion Matriks.

I. PENDAHULUAN

Pada saat ini banyak yang memakai atau menggunakan sosial media Twitter karena Twitter mudah digunakan oleh semua kalangan dari anak muda maupun orang tua. Karena dengan Twitter kita dengan mudah untuk memperoleh berbagai informasi maupun berkomunikasi dengan keluarga serta teman. Negara Indonesia merupakan negara yang terbesar kelima dalam penggunaan media Twitter dalam memberikan tweet setiap harinya.

Shopee pun telah membuat grup di Twitter untuk para konsumen-konsumennya dalam memberikan masukan-masukannya sehingga Shopee dapat memperbaiki kekurangannya dalam memberikan jasa kepada para konsumen-konsumennya. Shopee merupakan marketplace yang sering digunakan oleh masyarakat di Indonesia maupun masyarakat dari negara lain. Karena Shopee memberikan kemudahan untuk para konsumennya, seperti menjual barang-barang untuk kebutuhan sehari-hari maupun juga menjual tiket pesawat, tiket kereta api, pulsa, pemesanan hotel, dan masih banyak lagi.

Berdasarkan dari latar belakang yang telah dipaparkan di atas, maka peneliti membuat suatu penelitian yang berjudul "Sentimen Analisis Pelanggan Shopee di Twitter dengan Algoritma Naive Bayes". Penelitian dari Elly Indrayuni dan Mochamad Wahyudi (2015) yang melakukan penelitian berjudul "Penerapan Character N-Gram untuk Sentimen Analisis pada Riview Hotel dengan Algoritma Naive Bayes". Merupakan penelitian sebelumnya yang menjadi referensi utama untuk penelitian ini. Pada penelitian ini menggunakan penerapan N-Gram serta Naive Bayes. Penelitian ini mengambil data dari review-review hotel. Review tersebut dibagi menjadi 2 data yaitu review positif maupun review negatif. Perbedaan dari penelitian sebelumnya yaitu terletak pada pengambilan data nya serta penghitungannya. Pada penelitian ini menggunakan Rstudio untuk pengambilan data dari Twitter, serta untuk menghitung tingkat keakurasiannya menggunakan *Confusion Matriks*.

II. TINJAUAN PUSTAKA

Penelitian tentang sentimen analisis pada review pelanggan Tokopedia yang dilakukan [1] dengan menggunakan algoritma Naive Bayes, melakukan pengklasifikasian review-review yang ada pada ulasan-ulasan para pelanggan Tokopedia. Dengan menggunakan algoritma Naive Bayes dalam pengklasifikasian

riview-rivew terbukti bahwa riview-riview tersebut dapat dikategorikan menjadi dua kategori yaitu riview positif ataupun riview negatif.

Penelitian tentang penerapan character N-Gram untuk sentimen analisis pada riview hotel dengan algoritma Naive Bayes yang dilakukan [2]. Dalam penelitian ini mengambil data-data dari riview-riview pelanggan hotel. Dalam penelitian ini membuktikan bahwa dalam menerapkan character N-Gram pada proses preprocessing, dapat mempengaruhi tingkat akurasi yang dihasilkan dengan menggunakan klasifikasi Naive Bayes.

Penelitian tentang analisis sentimen dengan pelanggan JD.ID menggunakan metode Naive Bayes berbasis konversi ikon emosi yang dilakukan [3], penelitian ini mengambil data dari ulasan para pelanggan yang telah menggunakan jasa JD.ID dengan metode sampling dan Naive Bayes untuk melakukan suatu pengklasifikasian pada ulasan pelanggan JD.ID. Dengan proses pengklasifikasian dapat menghasilkan dua nilai yaitu nilai yang bersifat positif maupun nilai yang bersifat negatif. Dari penelitian ini menghasilkan bahwa tanpa pembobotan tf-idf dan konversi ikon emosi menghasilkan nilai akurasi 96,44% sedangkan dengan menggunakan pembobotan tf-idf dan konversi ikon emosi tingkat akurasinya meningkat yaitu 98%.

Penelitian tentang menganalisis sentimen pada opini film yang ada di Twitter yang dilakukan [4] melakukan pengambilan data dari hastag judul film yang ada pada sosial media Twitter. Dalam proses pengambilan datanya menggunakan API Twitter. Penelitian ini menghasilkan bahwa algoritma Naive Bayes dapat digunakan untuk pengklasifikasian data opini pada film berbahasa Indonesia.

Penelitian tentang analisis sentimen Twitter dengan Algoritma Naive Bayes yang dilakukan [5] dengan mengambil data dari Twitter berbahasa Indonesia sedangkan untuk proses pengklasifikasian menggunakan algoritma Naive Bayes. Algoritma Naive Bayes digunakan untuk melakukan sentimen analisis pada data yang telah diambil dari Twitter. Penelitian ini juga menggunakan confusion matriks untuk mengetahui berapa tingkat akurasi dalam menggunakan Naive Bayes untuk sentimen analisis. Setelah melakukan suatu proses penghitungan dapat menghasilkan tingkat akurasi sebesar 84 %.

Penelitian tentang menganalisis sentimen pada Twitter mahasiswa menggunakan algoritma backpropagation yang dilakukan [6] yang mengambil data responden mahasiswa dari grup mahasiswa yang ada di Twitter. Penelitian ini mempunyai tujuan untuk mengetahui kecenderungan emosi pada mahasiswa dengan sentimen analisis pada grup Twitter dengan metode backpropagation. Penelitian ini menghasilkan bahwa dengan metode propagation dapat mengetahui kecenderungan emosi pada mahasiswa.

Penelitian tentang analisis sentimen pada aplikasi Ruang Guru di Twitter menggunakan algoritma Klasifikasi yang dilakukan [7] melakukan penelitian dengan mengambil data-data dari komentar yang bersumber pada Twitter dengan

menggunakan metode Cross Industry Standard Process for Data Mining (CRISP-DM). Dalam penelitian ini menghasilkan bahwa menggunakan feature selection PSO dapat meningkatkan performa serta tingkat akurasi.

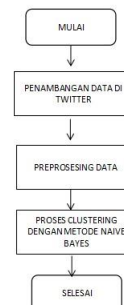
Penelitian tentang klasifikasi komentar pada twitter tentang pengesahan UUMD3 menggunakan metode KNN dan Naive Bayes yang dilakukan [8] mempunyai tujuan untuk mengimplementasikan algoritma KNN dengan algoritma Naive Bayes. Penelitian menghasilkan perbandingan antara data latih dengan data uji sebesar 66.7% untuk sedangkan 33.3% untuk data uji.

III. METODOLOGI PENELITIAN

Pada penelitian ini mengambil data-data dari Twitter yaitu di grup Shopee. Data yang diambil yaitu riview-riview dari para pemakai jasa Shopee. Riview dibagi menjadi 2 yaitu riview positif dan riview negatif. Jumlah data yang dipakai sebanyak 600 data terdiri dari 300 data yaitu 150 untuk komentar positif dan 150 untuk komentar negatif dan data latih 300 buah yang terbagi sama rata antara komentar positif dan negatif untuk data uji, dengan algoritma Naive Bayes. Pengambilan datanya menggunakan aplikasi Rstudio, setelah proses pengambilan datanya selesai dan sudah terbagi menjadi 2 data yaitu riview positif maupun negatif, maka dilakukan proses cleaning pada riview-riview yang telah diambil. Untuk pengujian tingkat akurasinya memakai algoritma Confusion Matriks.

A. Alur Penelitian

Diagram alur dalam penelitian



Gambar 1. Alur Penelitian

Pada diagram diatas merupakan proses alur dari penelitian. Dari alur tersebut dapat dijelaskan yaitu pertama kita mengambil data-data dari grup Shopee yang ada di Twitter, setelah kita mendapatkan datanya, kita melakukan preprocessing data yaitu cleaning, di proses ini kita membersihkan komentar-komentar, seperti menghilangkan hastag (#), titik (.), koma (,), ad (@), dan sebagainya. Setelah proses preprocessing data nya selesai kita masukan ke dalam aplikasi Naive Bayes. Di aplikasi Naive Bayes dilakukan proses stemming, serta melakukan penghitungan tingkat keakurasianya.

B. Naive Bayes

Algoritma Naive Bayes merupakan suatu metode classifier yang digunakan dalam suatu perhitungan probabilitas maupun statistik[1].

$$P(H|X) = \frac{P(X|H) \cdot P(H)}{P(X)} \quad (1)$$

Gambar 2. Rumus Naive Bayes

Keterangan :

- X** : Data yang belum diketahui
- H** : Hipotesis dari data class
- P(H|X)** : Probabilitas hipotesis nilai berdasarkan kondisi nilai X
- P(H)** : Probabilitas hipotesis nilai H
- P(X|H)** : Probabilitas nilai X berdasarkan hipotesis nilai H
- P(X)** : Probabilitas nilai X

C. R Studio

R Studio merupakan suatu aplikasi yang berfungsi untuk menulis program-program yang memakai bahasa R. RStudio merupakan aplikasi yang berbasis open source yang dapat digunakan di semua OS seperti Windows, Linux, dan Mac[9].

D. Convusion Matrix

Convusion Matrix merupakan alat yang digunakan untuk mengetahui suatu nilai dari tingkat keakurasian[5]. Convusion Matrix mempunyai 4 nilai yaitu True Positif (TP), False Positif (FP), False Negative (FN), dan True Negative (TN)[10]. Contoh tabel Convusion Matrix dapat dilihat pada gambar 2.

		Actual Values	
		1 (Positive)	0 (Negative)
Predicted Values	1 (Positive)	TP (True Positive)	FP (False Positive) Type I Error
	0 (Negative)	FN (False Negative) Type II Error	TN (True Negative)

Gambar 2. Tabel Convulsion Matriks

Keterangan :

- TP** : Pernyataan positif yang benar-benar terjadi
- TN** : Pernyataan negatif yang benar-benar terjadi
- FP** : Suatu prediksi positif, tetapi kebenarannya bersifat negatif
- FN** : Suatu prediksi negatif, tetapi kebenarannya bersifat positif

Cara Penghitungannya :

$$Accuracy = (TP + TN) / (TP + FP + FN + TN) \quad (2)$$

IV. HASIL DAN PEMBAHASAN

A. Tampilan RStudio

Ini merupakan tampilan dari software RStudio. Yang dipakai untuk memanggil rumus-rumus atau coding untuk mengambil data-data komentar-komentar dari grup Shopee yang ada di Twitter. Sebelum melakukan pengambilan data dengan Rstudio kita harus memasukan consumer key serta consumer secret nya yang ada pada Twitter kita, sehingga Rstudioa dapat mendeteksi Twitter kita. Setelah memasukan consumer key beserta consumer secretnya barulah kita bisa menjalankan rumus untuk pengambilan datanya.



Gambar 3. Tampilan RStudio

B. Rumus Untuk Mengambil Data dari Twitter

Ini adalah rumus dari aplikasi Rstudio untuk mengcrawling data dari Twitter. Data yang diambil adalah 100 data

1. Tw = searchTwitter('shopeeID', n = 100, retryOnRateLimit = 10e3) (4)

Rumus untuk menyimpan maupun membaca hasil data yang telah kita ambil ke komputer maupun laptop kita.

2. saveRDS(tw, file = 'tweet-mentah.rds') (5)

3. tw <- readRDS('tweet-mentah.rds') (6)

d = twListToDF(tw)

Kata yang diberi huruf tebal itu adalah nama yang akan kita simpan datanya.

C. Hasil Crawling Data

Ini merupakan gambar dari hasil kita mengcrawling data di Twitter. Disini terdapat banyak data yang telah kita ambil, tetapi data tersebut belum di cleaning masih belum teratur atau belum rapi.



Gambar 4. Hasil *Crawling* Data

D. Data Frame

Ini merupakan gambar dari hasil data frame yang telah kita ambil datanya dari grup Shopee yang ada di Twiter. Data tersebut belum dibagi menjadi riview positif maupun negatif, karena data tersebut masih campur antara data positif maupun negatif.

	text
1	semangat yaaa kak good luck hi
2	ashia mimin syatet deh nih kak cuss rintengin terus stingan a...
3	shla kakcp
4	wihh mantul banget nih kak ah
5	ntas deh kak mantul bangettt yukkk stay disini aja yaww jan...
6	min cek dm dong
7	aamin kak semangat trus ya buat ikutan komtisinya goodlu...
8	aamin makasih min
9	hi kak waah jangan simis dungg hehe slaa buat kejutan ha...
10	jangan simis dong kak harus usaha terus ah
11	aamin yah kak semangattttt av

Gambar 5. Hasil Data Frame

E. Program Naive Bayes

Ini adalah gambar tampilan dari program Naive Bayes. Pada tahap ini kita memproses data-data yang sudah diambil dari Twitter serta di Stemming. Dan proses penghitungan tingkat keakurasiannya.



Gambar 6. Tampilan Program Naive Bayes

F. Convusion Matrix

Setelah semua data-data komentar positif maupun negatif dari grup Shopee di Twitter dimasukkan ke program Naive Bayes, maka kita dapat melakukan penghitungan tingkat akurasi dengan menggunakan *Convusion Matrix* rumusnya yaitu :

$$\begin{aligned}
 \text{Accuracy} &= (TP+TN) / (TP+FP+FN+TN) \\
 &= (15+14) / (15+1+0+14) \\
 &= (29 / 30) * 100 \\
 &= 97\%
 \end{aligned}$$

V. KESIMPULAN DAN SARAN

Berdasarkan penelitian dan pengujian, maka dapat diambil kesimpulan sebagai berikut :

1. Tingkat keakurasi dalam *Sentimen Analysis* data dari grup Shopee di *Tweter* dalam menggunakan program *Naive Bayes* adalah sebesar 97%
2. Menghitung sentimen analisis dengan menggunakan metode Naive bayes sudah berhasil.
3. Dari hasil proses pengambilan data menggunakan Rstudio menghasilkan data positif dengan jumlah 150 data serta 150 data untuk data negatif.

Saran yang diberikan kepada pengembang penelitian untuk penelitian yang lebih lanjut yaitu ditambahkan perbandingan dari metode satu dengan metode lainnya agar hasil yang didapat lebih baik lagi,

UCAPAN TERIMA KASIH

Saya berterima kasih kepada Tuhan Yang Maha Esa yang telah memberi kesehatan pada saya ketika saya mengerjakan penelitian ini serta dosen pembimbing yang selama ini membimbing penelitian yang saya kerjakan. Orang tua saya yang membantu saya dan mendukung saya. Teman-teman saya yang telah membantu saya dalam penelitian yang saya kerjakan.

REFERENSI

- [1] Ai Nurhayatul Kamilah, 2017, *Analisa Sentimen Pelanggan Tokopedia Menggunakan Algoritma Naive Bayes Berdasarkan Riview Pelanggan*. Simki Techsain. Vol. 01 No. 06 Tahun 2017 ISSN : XXXX XXXX
- [2] Elly Indrayuni dan Mochamad Wahyudi. 2015, Penerapan Character N-Gram untuk Sentimen Analysis Riview Hotel Menggunakan Algoritma Naive Bayes. Konferensi Nasional Ilmu Pengetahuan dan Teknologi (KNIPT) 2015.
- [3] Fransiska Vina Sari dan Arief Wibowo, 2019, *Analisis Sentimen Pelanggan Toko Online JD.ID Menggunakan Metode Naive Bayes Classifier Berbasis Konversi Ikon Emosi* . Jurnal SIMETRIS, Vol. 10 No. 2 November 2019. ISSN : 2252-4983
- [4] Fajar Ratnawati, 2018, *Implementasi Algoritma Naive Bayes Terhadap Analisis Sentimen Opini Film Pada Twitter*. Jurnal INOVTEK Polbeng, Vol. 3 No 1 Juni 2018
- [5] Adhi Viky Sudiantoro dan Ery Zuliarso, 2018, *Analisis Sentimen Twitter Menggunakan Teks Mining Dengan Algoritma Naive Bayes Classifier*. Prosiding SINTAK 2018
- [6] Robet Habibi, Djoko Budiyanto Setyohadi, dan Ernawati, 2016, *Analisis Sentimen Pada Twitter Mahasiswa Menggunakan Metode BackPropagation*. INFORMATIKAN, Vol. 12 No. 1 April 2016
- [7] Angelina Puput Giovani , Ardiansyah, Tuti Haryanti, Laela Kurniawati, dan Windu Gata, 2020, *Analisis Sentimen Aplikasi Ruang Guru di Twitter Menggunakan Algoritma Klasifikasi*. Jurnal TEKNOINFO, Vol. 14 No. 2 2020. ISSN : 2615-224X
- [8] Putri Rizqiyah, 2018, *Klasifikasi Komentar Twitter Tentang Pengesahan UUMD3 Menggunakan Metode K_Nearest Neighbor (KNN) Dan Naive Bayes*.
- [9] Weksi Budiaji, 2019, *Penerapan Reproducible Research pada RStudio Dengan Bahasa R dan Paket Knitr*. Jurnal Ilmu Komputer dan Informatika, Vol. 5 No. 1 Juni 2019. ISSN: 2621-038X
- [10] Mercury Fluorida Fibrianda dan Adhitya Bhawiyuga, 2018, *Analisis Perbandingan Akurasi Deteksi Serangan Pada jaringan Komputer Dengan Metode Naive Bayes Dan Support Vector Machine (SVM)*. Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer, Vol. 2 No. 9, September 2018. ISSN: 2548-964X