

Implementasi Algoritma Naïve Bayes dalam Menganalisis Sentimen Review Pengguna Tokopedia pada Produk Kesehatan

Andi Ernawati, Ayu Ofta Sari, Siti Nurhaliza Sofyan, Muhammad Iqbal, Rian Farta Wijaya

Fakultas Ilmu Komputer Dan Teknologi Informasi, Universitas Pembangunan Panca Budi Medan, Medan, Indonesia

Email: aernawati296@gmail.com, ayuofta2849@gmail.com, sitnurhalizas102@gmail.com, muhammadiqbal@dosen.pancabudi.ac.id, rianfartawijaya@dosen.pancabudi.ac.id

Email Penulis Korespondensi: aernawati296@gmail.com

Abstrak– Harus di sadari bahwa Kepuasan pelanggan tujuan utama bagi perusahaan dalam mengembangkan usahanya. Karena opini para pelanggan yang dituliskan di media sosial akan memberikan pengaruh besar terhadap perusahaan maupun calon pelanggan. Dalam perkembangannya semakin banyak di temukan di berbagai media online salah satunya adalah tokopedia. Review produk merupakan salah satu sumber informasi yang penting terkait kualitas, pelayanan dan pengiriman baik dari konsumen maupun produsen. Dengan jumlah data yang sangat besar untuk setiap produk di tokopedia, dalam menganalisis dan menyimpulkan sebuah informasi ulasan produk pasti akan memakan banyak waktu jika dilakukan secara manual. Untuk mengatasi hal tersebut, diperlukan sistem analisis sentimen yang dapat secara otomatis mengekstraksi informasi penting yang secara objektif dapat menentukan kualitas produk dan menangani informasi tekstual yang besar. Sistem analisis sentimen terdiri dari beberapa tahapan, yaitu crawling, pre-processing, pembobotan kata, dan klasifikasi sentimen. Dengan menerapkan algoritma Naïve Bayes melalui pemilihan fitur range dan frekuensi, maka akan diperoleh hasil akurasi, akurasi dan recall dengan menggunakan uji Confusion Matrix. Dataset yang digunakan adalah dari situs kaggle.com tentang sentimen pelanggan pada produk kesehatan dengan jenis masker. menggunakan Metode Algoritma Naïve Bayes untuk mengetahui sentimen ulasan pengguna dengan klasifikasi 2 kelas positif dan negative dengan pendekatan NLP menghasilkan nilai akurasi sebesar 88%.

Kata Kunci: Algoritma Naïve Bayes; Sentimen; Tokopedia

Abstract – It must be realized that customer satisfaction is the main goal for companies in developing their business. Because customers' opinions written on social media will have a big influence on the company and potential customers. In its development, it is increasingly found in various online media, one of which is Tokopedia. Product reviews are an important source of information regarding quality, service and delivery from both consumers and manufacturers. With a very large amount of data for each product on Tokopedia, analyzing and concluding product review information will definitely take a lot of time if done manually. To overcome this, a sentiment analysis system is needed that can automatically extract important information that can objectively determine product quality and handle large amounts of textual information. The sentiment analysis system consists of several stages, namely crawling, pre-processing, word weighting, and sentiment classification. By applying the Naïve Bayes algorithm through selecting range and frequency features, accuracy, accuracy and recall results will be obtained using the Confusion Matrix test. The dataset used is from the kaggle.com site regarding customer sentiment on health products with the type of mask. using the Naïve Bayes Algorithm Method to determine the sentiment of user reviews by classifying 2 positive and negative classes using the NLP approach produces an accuracy value of 88%.

Keywords: Naïve Bayes Algorithm; Sentiment; Tokopedia

1. PENDAHULUAN

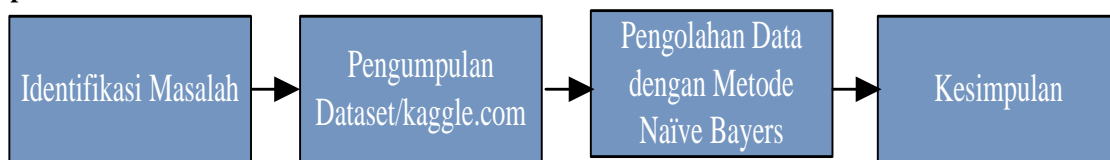
Perkembangan Penjualan online saat ini sangatlah pesat dikarenakan transaksi semakin mudah hanya dengan dari genggam tangan sudah bisa mendapatkan produk yang kita inginkan. Dengan demikian banyak juga permasalahan-permasalahan yang dihadapi yaitu dari pelayanan, jenis produk dan kualitas produk sehingga banyaknya pelanggan yang kecewa dengan kondisi seperti ini. Tujuan penelitian ini adalah menerapkan algoritma naïve bayes sebagai salah satu cara untuk mengatasi masalah yang terjadi. Dengan algoritma ini nantinya para pelanggan akan mengetahui produk mana yang baik pelayanan, jenis dan kualitas produknya karena dapat dilihat dari banyaknya sentimen para pelanggan dalam mereview produk yang sudah dibeli. Produk yang di bahas dipenelitian ini adalah produk kesehatan berjenis masker. Analisis ulasan produk dapat dengan mudah dilakukan dengan melihat peringkat bintang yang diberikan oleh pembeli, tetapi peringkat bintang tidak dapat mewakili isi keseluruhan ulasan. [1]. Analisis sentimen diaplikasikan untuk mengelompokkan komentar positif dan negatif dari konsumen sehingga mempercepat dan mempermudah tugas perusahaan untuk meninjau kembali kekurangan produk mereka Apabila ditemukan adanya sentimen negatif, maka perusahaan dapat dengan cepat mengambil tindakan untuk menanggulangnya. Sistem analisis sentimen yang akan dibangun pada penelitian ini menggunakan algoritma klasifikasi Naive Bayes. Ciri utama dari algoritma Naive Bayes adalah asumsi yang sangat kuat (naif) akan independensi dari masing- masing kondisi atau kejadian. Kelebihan dari Naive Bayes adalah proses klasifikasi data dapat disesuaikan dengan sifat dan kebutuhan masing-masing Dengan adanya sistem analisis sentimen ini diharapkan dapat masyarakat yang mempromosikan suatu produk pada marketplace mengetahui umpan balik terhadap produknya dan mengetahui opini yang disampaikan oleh pengguna marketplace terhadap produk yang ditawarkan.[2]

Penelitian yang dilakukan oleh Sulindawaty dkk tahun 2023. Tentang Penerapan Algoritma Naïve Bayes dalam Menganalisis Sentimen pada Review Pengguna E-Commerce. Sistem analisis sentimen review pengguna Shopee menggunakan metode Naïve Bayes ini menghasilkan nilai akurasi dengan satu kali pengujian dengan nilai sebesar accuracy sebesar 99,5%, precision sebesar 99,49%, recall sebesar 100%. Dengan demikian dapat disimpulkan metode klasifikasi algoritma Naïve Bayes cukup relevan meskipun akurasi nya belum 100% [2]. penelitian yang dilakukan Saiful Ulya dkk tahun 2022. Dengan judul Text Mining Sentimen Analisis Pengguna Aplikasi Marketplace Tokopedia Berdasar

Rating Dan Komentar Pada Google Play Store dengan hasil Pemodelan yang pertama dilakukan dengan Naïve Bayes dengan hasil akurasi mencapai 89.13%, kemudian pengujian kedua dengan Decision Tree dengan akurasi mencapai 83.51%, dan pengujian ketiga dengan Deep Learning dengan akurasi mencapai 84.91%.[3]. Penelitian yang dilakukan Bintang Zulfikar Ramadhan dkk tahun 2022 dengan judul Analisis Sentimen Ulasan Pada Aplikasi E-Commerce Dengan Menggunakan Algoritma Naïve Bayes dengan hasil Hasil evaluasi algoritma Naïve Bayes dengan tiga skenario berbeda dalam klasifikasi sentimen ulasan aplikasi E-Commerce menghasilkan hasil terbaik pada skenario 1 pada aplikasi Shopee (80% data training dan 20% data testing) yang menghasilkan accuracy 92%, precision 92,13%, recall 98,8% dan f1-score 95,35%.[4]. Penelitian yang dilakukan Ryfan Maulana Putra Hertaryawan dkk tahun 2023 Komparasi Algoritma Naive Bayes Dan Knearest Neighbor Pada Analisis Sentimen Terhadap Ulasan Pengguna Aplikasi Tokopedia dengan hasil Nilai akurasi Naive Bayes tanpa menggunakan fitur seleksi Particle Swarm Optimization adalah 76.30% dan nilai AUC 0.500, sedangkan Naive Bayes dengan menggunakan fitur seleksi Particle Swarm Optimization memiliki akurasi sebesar 74.09% dan nilai AUC 0.511, selisih peningkatan akurasi tanpa dan dengan menggunakan fitur seleksi adalah 4.29% dan selisih peningkatan AUC 0.011. [5]. Penelitian yang dilakukan Cahya Dewi dkk tahun 2022 Kepuasan Pengguna Layanan Shopee Food Menggunakan Algoritma Naive Bayes dengan hasil adanya pengaruh secara simultan antara tampilan aplikasi, jumlah driver, potongan harga dan metode pembayaran sebesar 0,497 terhadap kepuasan pengguna. Hal tersebut menunjukkan bahwa kemampuan faktor-faktor independen dalam penelitian ini mempengaruhi variabel dependen sebesar 49,7%, sedangkan sisanya 50,3% dijelaskan oleh variabel lain.[6]

2. METODOLOGI PENELITIAN

2.1. Tahapan Penelitian



Gambar 2.1. Tahapan Penelitian

1. Identifikasi Masalah

Identifikasi masalah dilakukan untuk menjelaskan dan mendefinisikan masalah-masalah yang dihadapi. Pada tahap ini juga dicari solusi atas masalah tersebut.

2. Pengumpulan Data

Pengumpulan data dilakukan dengan cara mengambil dataset dari kaggle.com

3. Pengolahan data dengan metode Naïve Bayers

Pengolahan data ini dilakukan disaat data sudah terkumpul dan dilakukan pengujian dengan metode naive bayers

4. Kesimpulan

Meyimpulkan hasil dari penelitian.

2.2 Algoritma

Algoritma adalah urutan langkah-langkah logis penyelesaian masalah yang disusun secara sistematis dan logis". Kata Logis adalah kunci dari algoritma. Langkah-langkah didalam algoritma wajib/harus logis dan dapat ditentukan nilainya benar dan salah. Dalam beberapa konteks algoritma adalah urutan langkah-langkah yang spesifikasi dalam melaksanakan pekerjaan tertentu[7]-[8]-[9]-

Algoritma adalah langkah - langkah dalam menyelesaikan masalah, sedangkan program adalah realisasi dari algoritma dalam bahasa pemrograman. Program ditulis dalam bahasa pemrograman dan kegiatan membuat program disebut pemrograman (programming). Orang/user yang menulis coding/program disebut pemrogram (programmer). Setiap langkah pada program disebut instruksi. Instruksi yang menjadi penghubung dari kode dan sistem ini yang bertanggung jawab untuk memberikan informasi yang tepat untuk dieksekusi.

Menurut para ahli definisi algoritma adalah yaitu: Algoritma adalah urutan langkah-langkah logis penyelesaian masalah yang disusun secara sistematis dan logis". Kata logis merupakan kata kunci dalam algoritma. Langkah-langkah dalam algoritma harus logis dan harus dapat ditentukan bernilai salah atau benar. Dalam beberapa konteks, algoritma adalah spesifikasi urutan langkah untuk melakukan pekerjaan tertentu[10]-[11]

2.3 Pengertian Naive Bayes

Bayes adalah klasifikasi menggunakan metode probabilistik dan statistik yang diusulkan oleh ilmuwan Inggris Thomas Bayes Dengan kata lain, memprediksi masa depan berdasarkan data masa lalu Naive Bayes menghitung, untuk setiap kelas keputusan, probabilitas bahwa kelas keputusan itu benar jika diberikan vektor informasi objek. Algoritma ini mengasumsikan bahwa atribut objek adalah independen Probabilitas dalam perkiraan akhir dihitung sebagai jumlah frekuensi dari tabel keputusan "master. [12]-[13]-[14]-[15]-[16]

2.4 Persamaan Metode Naive Bayes

Persamaan dari teorema Bayes adalah : [17]

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

Di mana :

X : Data dengan class yang belum diketahui

H : Hipotesis data merupakan suatu class spesifik

P(H|X) : Probabilitas hipotesis H berdasar kondisi X (posteriori probabilitas)

P(H) : Probabilitas hipotesis H (prior probabilitas)

P(X|H) : Probabilitas X berdasarkan kondisi pada hipotesis H

P(X) : Probabilitas X

Untuk menjelaskan metode Naive Bayes, perlu diketahui bahwa proses klasifikasi memerlukan sejumlah petunjuk untuk menentukan kelas apa yang cocok bagi sampel yang dianalisis tersebut. Karena itu, metode Naive Bayes di atas disesuaikan sebagai berikut:

$$P(C|F1 \dots Fn) = \frac{P(C)P(F1 \dots Fn|C)}{P(F1 \dots Fn)} \quad (2)$$

Di mana Variabel C merepresentasikan kelas, sementara variabel F1 ... Fn merepresentasikan karakteristik petunjuk yang dibutuhkan untuk melakukan klasifikasi. Maka rumus tersebut menjelaskan bahwa peluang masuknya sampel karakteristik tertentu dalam kelas C (Posterior) adalah peluang munculnya kelas C (sebelum masuknya sampel tersebut, seringkali disebut prior), dikali dengan peluang kemunculan karakteristik-karakteristik sampel pada kelas C (disebut juga likelihood), dibagi dengan peluang kemunculan karakteristik-karakteristik sampel secara global (disebut juga evidence). Karena itu, rumus di atas dapat pula ditulis secara sederhana sebagai berikut:

$$Posterior = \frac{prior \times likelihood}{evidence} \quad (3)$$

Nilai Evidence selalu tetap untuk setiap kelas pada satu sampel. Nilai dari posterior tersebut nantinya akan dibandingkan dengan nilai-nilai posterior kelas lainnya untuk menentukan ke kelas apa suatu sampel akan diklasifikasikan. Penjabaran lebih lanjut rumus Bayes tersebut dilakukan dengan menjabarkan $(C|F1, \dots, Fn)$ menggunakan aturan perkalian sebagai berikut:

$$P(C|F1, \dots, Fn) = P(C)P(F1, \dots, Fn|C)$$

$$= P(C)P(F1|C)P(F2, \dots, Fn|C, F1)$$

$$= P(C)P(F1|C)P(F2|C, F1)P(F3, \dots, Fn|C, F1, F2)$$

$$= P(C)P(F1|C)P(F2|C, F1)P(F3|C, F1, F2)P(F4, \dots, Fn|C, F1, F2, F3)$$

$$= P(C)P(F1|C)P(F2|C, F1)P(F3|C, F1, F2) \dots P(Fn|C, F1, F2, F3, \dots, Fn-1) \quad (4)$$

Dapat dilihat bahwa hasil penjabaran tersebut menyebabkan semakin banyak dan semakin kompleksnya faktor - faktor syarat yang mempengaruhi nilai probabilitas, yang hampir mustahil untuk dianalisa satu persatu. Akibatnya, perhitungan tersebut menjadi sulit untuk dilakukan. Di sinilah digunakan asumsi independensi yang sangat tinggi (naif), bahwa masing-masing petunjuk (F1, F2...Fn) saling bebas (independen) satu sama lain. Dengan asumsi tersebut, maka berlaku suatu kesamaan sebagai berikut:

$$P(F_i|F_j) = \frac{P(F_i \cap F_j)}{P(F_j)} = \frac{P(F_i)P(F_j)}{P(F_j)} = P(F_i) \quad (5)$$

Untuk $i \neq j$, sehingga

$$P(F_i|C, F_j) = P(F_i|C) \quad (6)$$

Persamaan di atas merupakan model dari teorema Naive Bayes yang selanjutnya akan digunakan dalam proses klasifikasi. Untuk klasifikasi dengan data kontinyu digunakan rumus Densitas Gauss :

$$P(X_i = x_i|Y = y_j) = \frac{1}{\sqrt{2\pi\sigma_{ij}}} e^{-\frac{(x_i - \mu_{ij})^2}{2\sigma_{ij}^2}} \quad (7)$$

Di mana :

P : Peluang

X_i : Atribut ke i x_i : Nilai atribut ke i Y : Kelas yang dicari y_i : Sub kelas Y

yang dicari μ : mean, menyatakan rata - rata dari seluruh atribut σ : Deviasi standar, menyatakan varian dari seluruh atribut.

3. HASIL DAN PEMBAHASAN

3.1. Data Alternatif

Dalam penelitian ini pengambilan data di lakukan dengan mengambil data set dari <https://www.kaggle.com/> sekitar 300 data yang akan di olah tentang produk kesehatan dengan jenis masker. Informasi yang didapat dalam format .csv

Tabel 1. Import dataset yang telah dilakukan text preprocessing

No	Nama_Produk	Akun	Ulasan	rate
1	Masker KF94 4 Ply Isi 10 Pcs Masker Medis Kese...	Stephanus	keduakali beli tapi sayang yang kedua lambat d...	3
2	Masker KF94 4 Ply Isi 10 Pcs Masker Medis Kese...	Yati	bagus pengiriman cepat	5
3	Masker KF94 4 Ply Isi 10 Pcs Masker Medis Kese...	Husin	barang sesuai pesanan	4
4	Masker KF94 4 Ply Isi 10 Pcs Masker Medis Kese...	Febby	agak panas pas dipakai	5
5	Masker KF94 4 Ply Isi 10 Pcs Masker Medis Kese...	Rahmat	barang sesuai permintaan dan bagus namun komun...	3

Pembagian data menjadi data sentimen berlabel positif dan negatif dengan angka 1 untuk positif dan angka 0 untuk negatif. Pengklasifikasian ini dilakukan pada ulasan yang memiliki rating 4 dan 5 sebagai sentimen positif dan rating 3 sampai 1 sebagai sentimen negatif.

Tabel 2. Klasifikasi data

	Nama_Produk	Akun	Ulasan	rate	label
1	Masker KF94 4 Ply Isi 10 Pcs Masker Medis Kese...	Stephanus	keduakali beli tapi sayang yang kedua lambat d...	3	Negatif
2	Masker KF94 4 Ply Isi 10 Pcs Masker Medis Kese...	Yati	bagus pengiriman cepat	5	Fositif
3	Masker KF94 4 Ply Isi 10 Pcs Masker Medis Kese...	Husin	barang sesuai pesanan	4	Fositif
4	Masker KF94 4 Ply Isi 10 Pcs Masker Medis Kese...	Febby	agak panas pas dipakai	5	Fositif
5	Masker KF94 4 Ply Isi 10 Pcs Masker Medis Kese...	Rahmat	barang sesuai permintaan dan bagus namun komun...	3	Negatif
...
29	Buy 1 Get 1 - Masker Primero - Sachtet	a***a	barang yg dikirim tdk sesuai deskripsi tidak t...	1	Negatif
29	Buy 1 Get 1 - Masker Primero - Sachtet	Teguh	kotaknya penyok disana sini mngkn dr pihak exp...	5	Fositif
29	Buy 1 Get 1 - Masker Primero - Sachtet	F***a	barang sesuai dan kualitas baik cuma kardusnya...	5	Fositif
29	Buy 1 Get 1 - Masker Primero - Sachtet	N***a	dapat harga promo jadi murah pengiriman agak l...	5	Fositif
30	Buy 1 Get 1 - Masker Primero - Sachtet	G***g	agak lama emang diproses sampe dikirim dan nya...	5	Fositif

300 rows × 10 columns

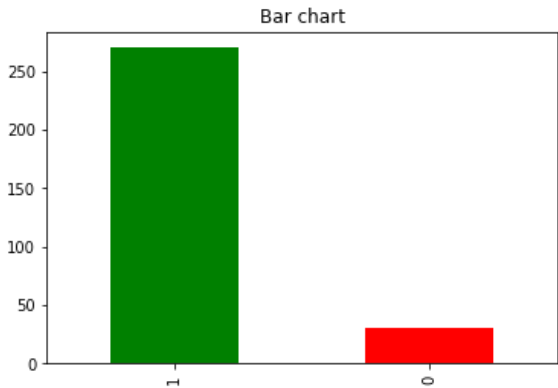
3.2 Pembobotan data

Setelah melakukan tahapan preprocessing data, data yang dihasilkan kemudian dilakukan tahap selanjutnya yaitu memberikan pembobotan data pada setiap kata dengan menggunakan metode TF-IDF dan selanjutnya melakukan klasifikasi menggunakan algoritma Naïve Bayes dan SVM seperti berikut..

[6.01396308 4.22220361 5.60849798 6.01396308 5.60849798 6.01396308
6.01396308 6.01396308 4.06805294 6.01396308 6.01396308 6.01396308
5.3208159 6.01396308 6.01396308 6.01396308 5.60849798 6.01396308
6.01396308 5.09767235 6.01396308 6.01396308 6.01396308 6.01396308
2.57997588 6.01396308 5.60849798 6.01396308 6.01396308 6.01396308
4.9153508 5.3208159 6.01396308 5.3208159 6.01396308 3.61606781
6.01396308 6.01396308 2.36330484 5.60849798 6.01396308 6.01396308
5.3208159 5.09767235 3.27312306 6.01396308 6.01396308 6.01396308
5.3208159 6.01396308 6.01396308 6.01396308 6.01396308 6.01396308
6.01396308 4.62766872 6.01396308 6.01396308 4.76120012 6.01396308
6.01396308 5.60849798 6.01396308 4.76120012 6.01396308 5.3208159
6.01396308 6.01396308 5.09767235 6.01396308 6.01396308 2.47300376
6.01396308 6.01396308 6.01396308 6.01396308 6.01396308 6.01396308
6.01396308 6.01396308 5.60849798 5.60849798 6.01396308 5.60849798
6.01396308 5.3208159 6.01396308 6.01396308 6.01396308 5.60849798
6.01396308 6.01396308 5.60849798 5.09767235 5.60849798 4.22220361
6.01396308 6.01396308 6.01396308 6.01396308 6.01396308 6.01396308
5.60849798 6.01396308 6.01396308 6.01396308 6.01396308 6.01396308
6.01396308 5.60849798 4.62766872 6.01396308 6.01396308 6.01396308
5.3208159 5.09767235 6.01396308 5.60849798 6.01396308 6.01396308
6.01396308 6.01396308 4.40452517 6.01396308 6.01396308 6.01396308
6.01396308 5.09767235 6.01396308 5.09767235 5.3208159 4.76120012
6.01396308 4.76120012 6.01396308 6.01396308 6.01396308 3.93452154
6.01396308 5.3208159 6.01396308 6.01396308 5.3208159 6.01396308
5.60849798 5.60849798 6.01396308 6.01396308 6.01396308 6.01396308
6.01396308 6.01396308 5.3208159 6.01396308 6.01396308 6.01396308
6.01396308 5.60849798 5.60849798 6.01396308 5.60849798 5.60849798
6.01396308 6.01396308 6.01396308 5.60849798 5.60849798 6.01396308
6.01396308 6.01396308 6.01396308 4.40452517 6.01396308 6.01396308
6.01396308 6.01396308 4.40452517 6.01396308 6.01396308 4.9153508
5.3208159 5.3208159 3.52905643 5.60849798 6.01396308 6.01396308
6.01396308 6.01396308 6.01396308 6.01396308 4.76120012 6.01396308
6.01396308 5.60849798 5.60849798 6.01396308 6.01396308 5.3208159
6.01396308 6.01396308 2.71812622 6.01396308 4.22220361 5.60849798
5.60849798 5.60849798 5.3208159 5.09767235 6.01396308 3.57161605
6.01396308 6.01396308 4.9153508 6.01396308 6.01396308 6.01396308
5.60849798 6.01396308 4.9153508 5.09767235 6.01396308 6.01396308
6.01396308 4.76120012 5.3208159 6.01396308 6.01396308 6.01396308
6.01396308 6.01396308 4.50988569 6.01396308 5.60849798 6.01396308
6.01396308 5.60849798 5.3208159 6.01396308 6.01396308 5.3208159
6.01396308 6.01396308 3.76267129 6.01396308 6.01396308 6.01396308
5.3208159 6.01396308 5.60849798 6.01396308 6.01396308 6.01396308
2.92292063 5.3208159 5.3208159 6.01396308 6.01396308 5.09767235
6.01396308 6.01396308 6.01396308 6.01396308 6.01396308 5.60849798
6.01396308 6.01396308 6.01396308 6.01396308 4.50988569 6.01396308
5.60849798 5.60849798 6.01396308 4.62766872 6.01396308 6.01396308
6.01396308 5.60849798 6.01396308 6.01396308 5.60849798 6.01396308
6.01396308 6.01396308 6.01396308 6.01396308 6.01396308 3.71137799
4.9153508 6.01396308 6.01396308 5.60849798 3.99906006 4.62766872
6.01396308 6.01396308 6.01396308 6.01396308 6.01396308 6.01396308
5.60849798 6.01396308 4.06805294 5.09767235 6.01396308 6.01396308
6.01396308 6.01396308 5.3208159 5.60849798 3.57161605 5.60849798
5.3208159 5.09767235 3.66258783 6.01396308 5.09767235 5.60849798
6.01396308 6.01396308 5.09767235 6.01396308 5.60849798 6.01396308
5.3208159 4.40452517 5.3208159 3.30591288 6.01396308 6.01396308
6.01396308 6.01396308 5.60849798 6.01396308 4.50988569 5.09767235
5.09767235 5.60849798 6.01396308 6.01396308 5.60849798 6.01396308

3.93452154 5.60849798 6.01396308 4.14216091 6.01396308 4.62766872
 6.01396308 6.01396308 6.01396308 5.3208159 5.09767235 6.01396308
 5.60849798 6.01396308 6.01396308 4.62766872 6.01396308 6.01396308
 3.57161605 6.01396308 6.01396308 5.60849798 4.62766872 5.60849798
 6.01396308 6.01396308 6.01396308 3.66258783 6.01396308 6.01396308
 5.60849798 5.3208159 5.3208159 6.01396308 6.01396308 6.01396308
 6.01396308 6.01396308 6.01396308 6.01396308 5.09767235 6.01396308
 6.01396308 6.01396308 4.62766872 5.60849798 5.09767235 6.01396308
 6.01396308 5.60849798 6.01396308 5.60849798 6.01396308 6.01396308
 4.06805294 6.01396308 6.01396308 6.01396308 5.60849798 2.45861502
 6.01396308 5.09767235 6.01396308 5.3208159 6.01396308 6.01396308
 5.60849798 6.01396308 6.01396308 6.01396308 5.09767235 6.01396308
 6.01396308 6.01396308 5.60849798 6.01396308 4.9153508 6.01396308
 5.3208159 6.01396308 6.01396308 6.01396308 4.76120012 5.3208159
 6.01396308 6.01396308 6.01396308 5.60849798 5.60849798 5.3208159
 6.01396308 6.01396308 6.01396308 6.01396308 6.01396308 6.01396308
 3.4112734 5.3208159 5.3208159 4.50988569 5.60849798 6.01396308
 4.22220361 6.01396308 5.60849798 6.01396308 4.50988569 6.01396308
 6.01396308 4.50988569 6.01396308 6.01396308 5.3208159 6.01396308
 6.01396308 5.3208159 6.01396308 6.01396308 5.60849798 6.01396308
 6.01396308 6.01396308 5.3208159 6.01396308 5.60849798 6.01396308
 6.01396308 6.01396308 6.01396308 6.01396308 4.62766872 6.01396308
 6.01396308 4.9153508 4.14216091 6.01396308 5.60849798 6.01396308]
 (0, 412) 0.2467579753658206
 (0, 401) 0.11901174168830964
 (0, 386) 0.22400697531094216
 (0, 369) 0.5318737814049063
 (0, 313) 0.2467579753658206
 (0, 302) 0.196918208267884
 (0, 216) 0.2714850052676621
 (0, 200) 0.26314728606809085
 (0, 176) 0.21320548648766965
 (0, 157) 0.2714850052676621
 (0, 137) 0.19045448643279128
 (0, 109) 0.2714850052676621
 (0, 48) 0.25755946418909315
 (0, 44) 0.15843882534096299
 (0, 38) 0.11439815623698339
 (0, 24) 0.12488633646281205

Dapat dilihat bahwa isi ulasan produk lebih banyak pada label sentimen 1 atau ulasan dengan rating positif ini berarti pelanggan yang menggunakan marketplace Tokopedia dan melakukan transaksi pembelian pada produk masker Kesehatan merasa puas bertansaksi di marketplace Tokopedia dan prosuk masker Kesehatan sehingga memberikan feedback atau ulasan komentar lebih banyak yang positif.



Gambar 2. Ulasan Produk

Pada proses ini kami menggunakan library sklearn.model_selection dengan modul train_test_split untuk membagi data latih (X_train dan y_train) dan data uji (X_test dan y_test) dengan persentasi data latih 70% dan data uji 30% serta memilih

label data yaitu yang merupakan variable independen dari data kami yaitu kolom label untuk dijadikan parameter klasifikasi prediksi.

```

----- classification report -----
              precision    recall  f1-score   support

     0           0.50      0.33      0.40         3
     1           0.93      0.96      0.95        27

 accuracy               0.90         30
 macro avg              0.71      0.65      0.67         30
 weighted avg          0.89      0.90      0.89         30
  
```

Gambar 3. Laporan klasifikasi

3.3 WordCloud

Setelah mengetahui setiap kata, proses selanjutnya menampilkan visualisasi *WordCloud*. Terdapat pada Gambar dibawah ini..

Tabel 3. Ulasan Bernilai Negatif

No	Nama_Produk	Akun	Ulasan_clean	label
1	Masker KF94 4 Ply Isi 10 Pcs Masker Medis Kese...	Stephanus	dua kali beli sayang lambat respon order pagi ...	0
5	Masker KF94 4 Ply Isi 10 Pcs Masker Medis Kese...	Rahmat	barang sesuai minta bagus komunikasi jual beli...	0
9	Masker KF94 4 Ply Isi 10 Pcs Masker Medis Kese...	n***l	proses barang jemput kurir banget anjas pokok	0
14	MASKER 3PLY IZIN KEMENKES MASKER MEDIS KESEHAT...	Renold	kecewa berat masker hijab warna hitam tali mas...	0
16	MASKER 3PLY IZIN KEMENKES MASKER MEDIS KESEHAT...	Ridwan	karet headloop nya pendek istri pakai untung anak	0
36	Masker Evo PlusMed 4d Medis - Putih	J***a	pesan masker warna hijau datang warna pink	0
57	Disposable Mask KN95 N95 Facemask Kesehatan Ma...	Anggia	bungkus plastik sealed debu kotor	0
85	ONEMED - Masker Karet 3 Ply / Masker Non-Medis...	Adhi	tipis sih ragu banding type jenis ply ply	0
104	C care Duckbill 3D Stereo 3 Ply Isi 50 Pcs Mas...	Risca	masker tipis ya bunda	0
116	Masker KN95 Alkindo Masker Medis 5ply 1 Box 10...	Duarte	kualitas masker jelek masker bolong jahit isi ...	0
117	Masker KN95 Alkindo Masker Medis 5ply 1 Box 10...	M***a	size telinga	0
166	Facemask 50s / BOX Masker Wajah Kesehatan 3ply...	NURHASANAH	fast respon kirim cepat sayang box khusus hija...	0
170	Facemask 50s / BOX Masker Wajah Kesehatan 3ply...	Jhon	pesan box hijab box model hijab	0
171	Masker Karet Putih 3 Ply Daily Use Onemed Box ...	Luke	masuk debu masuk masker hari	0
173	Masker Karet Putih 3 Ply Daily Use Onemed Box ...	H***h	tipis banget masker beda onemed warna putih tipis	0

kesehatan), harga, kualitas pengiriman atau pengemasan, serta kualitas produk masker Kesehatan yang paling banyak disebutkan pelanggan dalam hasil Analisa ulasan sentimen yang negatif.

Tabel 4. Ulasan Bernilai Fositif

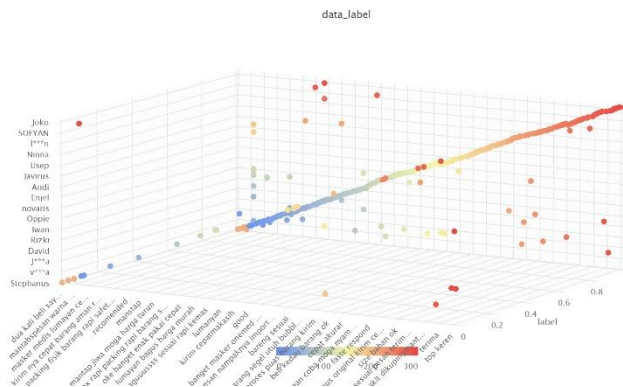
	Nama_Produk	Akun	Ulasan_clean	label
2	Masker KF94 4 Ply Isi 10 Pcs Masker Medis Kese...	Yati	bagus kirim cepat	1
3	Masker KF94 4 Ply Isi 10 Pcs Masker Medis Kese...	Husin	barang sesuai pesan	1
4	Masker KF94 4 Ply Isi 10 Pcs Masker Medis Kese...	Febby	panas pakai	1
6	Masker KF94 4 Ply Isi 10 Pcs Masker Medis Kese...	Alden	bagus masker	1
7	Masker KF94 4 Ply Isi 10 Pcs Masker Medis Kese...	Heni	pengirimn cepat produk bagus konsumen suka mak...	1
...
294	Buy 1 Get 1 - Masker Primero - Sachet	M***y	top keren	1
297	Buy 1 Get 1 - Masker Primero - Sachet	Teguh	kotak penyok sana mngkn expediai	1
298	Buy 1 Get 1 - Masker Primero - Sachet	F***a	barang sesuai kualitas kardus penyok next time...	1
299	Buy 1 Get 1 - Masker Primero - Sachet	N***a	harga promo murah kirim promo	1
300	Buy 1 Get 1 - Masker Primero - Sachet	G***g	proses kirim sabar ya temanteman beli paket cu...	1

270 rows x 4 columns



Gambar 5. Ulasan Fositif

Dari visualisasi diatas merupakan wordcloud kata yang paling banyak muncul pada ulasan yang memiliki label sentimen positif. Kata yang paling sering muncul dan mengarah ke ulasan positif membahas seputar : barang, cepat, bagus, masker, aman, kualitas, sesuai, rapi, respon, aman, recommended, dan sebagainya. Sehingga dari kata-kata ini bisa menjadi masukan untuk penjual dan marketplace Tokopedia untuk menjaga kualitas atau meningkatkan kembali kualitas barang (produk masker kesehatan), kualitas yang sesuai dan aman, serta respon penjual paling banyak disebutkan pelanggan dalam hasil Analisa ulasan sentimen yang positif.



Gambar 6. Pengujian Data dengan aplikasi Rapid Miner

3. KESIMPULAN

Dari hasil penelitian menggunakan Metode Algoritma Naïve Bayes untuk mengetahui sentimen ulasan pengguna dengan klasifikasi 2 kelas positif dan negative dengan pendekatan NLP menghasilkan nilai akurasi sebesar 88%. Selain itu, didapatkan bahwa Analisa Sentimen pada ulasan marketplace Tokopedia pada produk masker kesehatan menunjukkan lebih banyak pada ulasan yang positif. Ini berarti pelayanan dan produk masker Kesehatan yang disediakan di marketplace Tokopedia sudah cukup baik.

Dari hasil Analisis diatas dapat disimpulkan hasil scraping yang kami dapat dari produk pencarian masker kesehatan pada Tokopedia menampilkan data yang menunjukkan ulasan positif lebih dominan daripada hasil ulasan negatif dan untuk ulasan negatif kata yang paling sering muncul adalah seputar kualitas produk yang tipis, mudah putus, sobek atau bolong, kualitas karetnya dan pada pengemasan yaitu kotak/dus penyok dan untuk analisa positif kata yang paling sering muncul adalah seputar kualitas produk yang sesuai dan rapi, pengiriman yang aman dan cepat dan respon penjual..

REFERENCES

- [1] S. M. Salsabila, A. A. Murtopo, and N. Fadhilah, "Analisis Sentimen Pelanggan Tokopedia Menggunakan Metode Naïve Bayes Classifier," vol. 11, no. September, pp. 30–35, 2022.
- [2] E. Laia and M. Yamin, "Penerapan Algoritma Naïve Bayes dalam Menganalisis Sentimen pada Review Pengguna E-Commerce," vol. 4, no. 1, pp. 305–316, 2023, doi: 10.30865/klik.v4i1.1186.
- [3] P. A. D. A. G. Oogle and P. L. A. Y. S. Tore, "T Ext M Ining S Entimen A Nalisis P Engguna A Plikasi M Arketplace T Okopedia B Erdasar R Ating D An K Omentar," pp. 33–40, 2022.
- [4] B. Z. Ramadhan, I. Riza, and I. Maulana, "Analisis Sentimen Ulasan Pada Aplikasi E-Commerce Dengan Menggunakan Algoritma Naïve Bayes," vol. 6, no. 2, pp. 220–225, 2022.
- [5] R. Maulana, P. Hertaryawan, M. Raihan, and I. Santoso, "KOMPARASI ALGORITMA NAIVE BAYES DAN K- NEAREST NEIGHBOR PADA ANALISIS SENTIMEN," vol. 17, no. 2, pp. 177–189, 2023.
- [6] C. D. A. A. P. Chrishariyani *et al.*, "Kepuasan Pengguna Layanan Shopee Food Menggunakan Algoritma Naive Bayes," vol. 02, pp. 99–106, 2022, doi: 10.21456/vol12iss2pp99-106.
- [7] M. Bobbi, K. Nasution, A. Karim, and S. Esabella, "Sistem Pendukung Keputusan Penilaian Kinerja Ketua Program Studi Menerapkan Metode WASPAS dengan Pembobotan ROC," vol. 4, no. 1, pp. 130–136, 2022, doi: 10.47065/bits.v4i1.1619.
- [8] A. Karim, "Penerapan Algoritma Entropy dan Aras Menentukan Desa Terbaik Di Pemerintah Kabupaten Labuhanbatu," vol. 3, no. 1, pp. 33–43, 2022.
- [9] A. Karim, S. Esabella, T. Andriani, and M. Hidayatullah, "Penerapan Metode Multi-Objective Optimization on the Basis of Simple Ratio Analysis (MOOSRA) dalam Penentuan Lulusan Mahasiswa Terbaik," vol. 4, no. 1, pp. 162–168, 2022, doi: 10.47065/bits.v4i1.1630.
- [10] G. Ginting, S. Alvita, A. Karim, and M. Syahrizal, "Penerapan Complex Proportional Assessment (COPRAS) Dalam Penentuan Kepolisian Sektor Terbaik," vol. 4, no. 2, pp. 616–631, 2020.
- [11] A. Ernawati, "Penerapan Algoritma Entropy Dan Aras Menentukan Penerima Beasiswa Mahasiswa Berprestasi Di Pemerintah Kabupaten Labuhanbatu," vol. 3, no. 2, pp. 74–84, 2022.
- [12] K. Sari and R. P. Cahyono, "Analisis Mood Twitter Melalui Layanan J & T Express Selama Harbolnas 12 / 12 (Dengan Metode Naive Bayes)," vol. 2, no. 11, pp. 1–16, 2022.
- [13] V. No, A. A. Sari, M. K. Anam, and M. Jamaris, "Edumatic : Jurnal Pendidikan Informatika Sistem Prediksi Keuntungan Influencer Pengguna E-Commerce Shopee Affiliates menggunakan Metode Naïve Bayes," vol. 6, no. 2, pp. 394–403, 2022, doi: 10.29408/edumatic.v6i2.6787.
- [14] T. H. Pudjiantoro, F. R. Umbara, and B. Trihatmoko, "Analisis Sentimen Terhadap E-commerce Pada Media Sosial Twitter Menggunakan Metode Naïve bayes," pp. 1–7, 2021.
- [15] J. J. A. Limbong, I. Sembiring, K. D. Hartomo, U. Kristen, S. Wacana, and P. Korespondensi, "Analisis Klasifikasi Sentimen Ulasan Pada E-Commerce Shopee Berbasis Word Cloud Dengan Metode Naive Bayes Dan K-Nearest Analysis Of Review Sentiment Classification On E-Commerce Shopee Word Cloud Based With Naïve Bayes And K-Nearest Neighbor Methods," vol. 9, no. 2, pp. 347–356, 2022, doi: 10.25126/jtiik.202294960.
- [16] A. Kepuasan *et al.*, "Analisis kepuasan penggunaan aplikasi shopee menggunakan algoritma naïve bayes," pp. 3–6.

- [17] A. Saleh, "Implementasi Metode Klasifikasi Naïve Bayes Dalam Memprediksi Besarnya Penggunaan Listrik Rumah Tangga," vol. 2, no. 3, pp. 207–217, 2015.
- [18] M. Bobbi, K. Nasution, A. Karim, and S. Esabella, "Sistem Pendukung Keputusan Penilaian Kinerja Ketua Program Studi Menerapkan Metode WASPAS dengan Pembobotan ROC," vol. 4, no. 1, pp. 130–136, 2022, doi: 10.47065/bits.v4i1.1619.
- [19] H. Syahputra, M. Syahrizal, S. Suginam, S. D. Nasution, and B. Purba, "SPK Pemilihan Konten Youtube Layak Tonton Untuk Anak-Anak Menerapkan Metode Additive Ratio Assessment (ARAS)," Semin. Nas. Teknol. Komput. Sains, vol. 1, no. 1, pp. 678–685, 2019, [Online]. Available: <https://prosiding.seminar-id.com/index.php/sainteks/article/view/215/210>.
- [20] S. Alex Rizky Saputra, "Implementasi Algoritma ARAS Pada SPK untuk Menentukan Peringkat Dosen Terbaik," Indones. J. Comput. Sci., vol. 11, no. 1, pp. 578–591, 2022.
- [21] A. B. Ginting, "Implementasi Metode Additive Ratio Assessment (ARAS) Dalam Menentukan Perusahaan Penyalur Tenaga Kerja Terbaik," vol. 9, pp. 174–182, 2021.
- [22] A. Karim, S. Esabella, T. Andriani, and M. Hidayatullah, "Penerapan Metode Multi-Objective Optimization on the Basis of Simple Ratio Analysis (MOOSRA) dalam Penentuan Lulusan Mahasiswa Terbaik," vol. 4, no. 1, pp. 162–168, 2022, doi: 10.47065/bits.v4i1.1630.