

## **Penerapan Algoritma Naïve Bayes dalam Menganalisis Sentimen pada Review Pengguna E-Commerce**

**Julius Yoshua Siahaan<sup>1</sup>, Aries Saifudin<sup>2</sup>**

Teknik Informatika, Universitas Pamulang, Tangerang Selatan, Indonesia

E-mail: [juliusyosuasiahaan@outlook.com](mailto:juliusyosuasiahaan@outlook.com)

**Abstrak-** Dalam e-commerce Shopee, proses jual beli terus berjalan setiap harinya, dan komentar yang diberikan konsumen pun akan semakin banyak. Komentar yang diberikan konsumen akan menjadi referensi/review terhadap suatu produk yang telah dibeli oleh konsumen. Konsumen dengan leluasa memberikan ulasan berisi komentar positif dan komentar negatif pada kolom Komentar yang tertera pada website e-commerce Shopee. Dengan menerapkan algoritma Naïve Bayes melalui pemilihan fitur range dan frekuensi, maka akan diperoleh hasil akurasi, akurasi dan recall dengan menggunakan uji Confusion Matrix. API digunakan untuk mengumpulkan data dan melakukan simulasi menggunakan ulasan data di situs web Shopee untuk menghasilkan sentimen positif dan negatif. Pengujian akan menampilkan hasil nilai persentil untuk accuracy dan retrieval. Dari hasil pengujian diperoleh hasil accuracy sebesar 99,5%, precision sebesar 99,49%, recall sebesar 100%. Dengan demikian dapat disimpulkan metode klasifikasi algoritma Naïve Bayes cukup relevan meskipun akurasi belum 100%. Hasil penelitian ini dapat memberikan solusi bagi toko Shopee untuk meningkatkan mutu penjualannya.

**Kata Kunci:** Sentimen; Review; Shopee; Naïve Bayes; Shopee;

**Abstract-** In Shopee e-commerce, the buying and selling process continues every day, and the comments given by consumers will increase. Comments given by consumers will become a reference/review of a product that has been purchased by the consumer. Consumers are free to provide reviews containing positive and negative comments in the Comments column listed on the Shopee e-commerce website. 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. API is used to collect data and perform simulations using review data on the Shopee website to generate positive and negative sentiment. The test will display the results of the percentile values for accuracy and retrieval. From the test results, accuracy was 99.5%, precision was 99.49%, recall was 100%. Thus, it can be concluded that the Naïve Bayes algorithm classification method is quite relevant even though its accuracy is not 100%. The results of this research can provide solutions for Shopee stores to improve the quality of their sales.

**Keywords:** Sentiment; Review; Shopee; Naive Bayes; Shopee;

### **1. PENDAHULUAN**

Jual beli secara online telah menjadi bagian dari kebutuhan belanja masyarakat saat ini. Salah satu marketplace dengan model e-commerce yang populer di Indonesia adalah Shopee. Berdasarkan data Peta E-Commerce Indonesia yang dikeluarkan oleh iPrice, Shopee adalah salah satu e-commerce regional yang beroperasi di hampir seluruh negara di Asia Tenggara, menduduki peringkat pertama ranking aplikasi e-commerce pada Play Store maupun App Store dengan jumlah pengunjung bulanan sebanyak 134,383,300 pengunjung pada kuartal 3 2021 [1].

Shopee menawarkan pengguna dalam melakukan sebuah transaksi barang secara online. Namun begitu, tidak jarang konsumen merasa ragu untuk membeli produk secara online dengan harga pasar yang cukup tinggi seperti produk smartphone. Hal ini dibuktikan berdasarkan hasil penelitian sebelumnya yang mendapati sebuah kesimpulan bahwa kepercayaan pelanggan mempengaruhi niat pembelian produk smartphone di Shopee Indonesia. Pelanggan dinilai masih merasa kurang percaya terhadap para penjual smartphone dalam memenuhi keinginan para pelanggan [2].

Ulasan produk merupakan salah satu sumber informasi tentang kualitas produk dan sangat berpengaruh pada konsumen dan produsen. Dalam kegiatan pembelian barang di marketplace, pembeli dapat memberikan ulasan setelah menerima barang yang dibeli. Pelanggan atau klien yang merasa tidak puas dengan layanan atau produk yang ditawarkan biasanya akan menuliskan

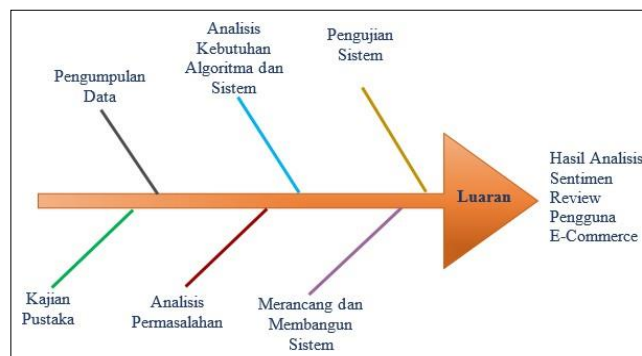
keluhannya di media sosial. Ulasan pembelian produk terdiri dari bintang dan isi komentar ulasan yang berisi tanggapan, apresiasi maupun kritik dan masukan pada produk yang telah dibeli tersebut [7]. Kepuasan pelanggan merupakan hal penting yang menjadi tujuan perusahaan. Disadari atau tidak, opini-opini pelanggan yang dituliskan di media sosial, sedikit atau pun banyak, akan memberikan pengaruh pada calon pelanggan. Dalam perkembangannya sentimen semakin mudah ditemukan dalam berbagai media online seperti shopee.

Pengertian review ini bisa positif dan negatif, menurut orang yang melakukan review. Namun sebagian besar review didasarkan pada informasi yang diperoleh sebelumnya dari produk sehingga dapat dilakukan dengan cara yang benar atau membawa hal negatif pada produk.

Opinion Mining adalah salah satu bidang ilmu komputer yang mempelajari komputasi linguistik, pemrosesan bahasa alami, dan penambahan teks yang bertujuan menganalisis emosi, penilaian, sikap, opini, sentimen, evaluasi seseorang terhadap pembicara atau penulis yang senang dengan suatu produk, layanan, organisasi, individu, tokoh masyarakat, topik, peristiwa, atau aktivitas tertentu [8][9].

## 2. METODOLOGI

Agar penelitian ini memperoleh hasil sesuai dengan tujuan yang diharapkan, maka dilakukan langkah-langkah sebagai berikut:



*Gambar 1. Tahapan Penelitian*

Berdasarkan gambar 1, tahapan penelitian analisis sentimen terhadap review pengguna e-commerce diuraikan sebagai berikut:

### a. Kajian Pustaka

Kajian pustaka yang ditelaah terkait dengan review produk, sistem analisis sentimen, algoritma Naïve Bayes dan konsep lainnya yang terkait dengan masalah penelitian ini. Kajian pustaka dilakukan dengan menelaah sumber-sumber yang bersumber dari jurnal-jurnal maupun hasil penelitian lainnya sebagai referensi penelitian terdahulu, buku teks, sumber online (internet) dan sumber lainnya yang terkait dengan topik penelitian yang dibahas.

### b. Pengumpulan Data

Pengumpulan data terkait review produk Iphone XR 256GB/128GB/64GB di toko Kinophonecell website Shopee. Data diambil menggunakan API yang telah disediakan oleh Shopee dan di crawling dengan menggunakan bahasa pemrograman PHP, kemudian data akan disimpan ke database.

c. Analisis Permasalahan

Menganalisis permasalahan terkait review produk Iphone XR 256GB/128GB/64GB oleh pengguna shopee di toko Kinophonecell.

d. Analisis Kebutuhan Algoritma dan Sistem

Menganalisis kebutuhan penerapan algoritma Naïve Bayes dan pembuatan sistem analisis sentimen yang akan digunakan untuk penyelesaian masalah pada penelitian ini.

e. Merancang dan Membangun Sistem Analisis Sentimen

Merancang dan Membangun sistem analisis sentimen dengan menerapkan algoritma naïve bayes ke dalam sistem untuk digunakan sebagai perhitungan dalam mengelompokkan sentimen ke dalam kriteria positif, dan negatif.

f. Pengujian Sistem

Dilakukan dengan menggabungkan modul-modul yang sudah dibuat dan dilakukan pengujian ini dilakukan untuk mengetahui apakah aplikasi yang dibuat telah sesuai dengan desain yang telah dirancang sebelumnya dan masih terdapat kesalahan atau tidak. Pengujian yang dilakukan berupa mencoba menguji sistem yang dihasilkan dengan menginputkan data review pelanggan untuk dilihat apakah sistem yang dihasilkan dapat mengelompokkan data review ke dalam kelompok positif, atau negatif.

### 3. HASIL DAN PEMBAHASAN

Dalam penelitian ini data yang digunakan adalah data sekunder, yaitu data yang diperoleh dalam penelitian ini berdasarkan komentar positif dan ulasan negatif terhadap produk berbahasa Indonesia di website e-commerce. Dataset yang digunakan dalam penelitian berjumlah 153 data komentar.

- A. Normalisasi Kasus adalah tahap mengubah kalimat menjadi huruf kecil. Pada tahap ini akan dilakukan pengecekan dataset per kalimat untuk mengubah kalimat menjadi huruf kecil.

Table 1. Contoh proses normalisasi kasus (lowercase)

No.	Data sebelum Proses Normalisasi Kasus	Data setelah Proses Normalisasi Kasus
1.	barang yang datang beda sekali sama yang di gambar.. customer amat sangat kecewa..	barang yang datang beda sekali sama yang di gambar.. customer amat sangat kecewa..

- B. Tokenisasi adalah suatu tahap pemecahan teks yang dapat berupa kalimat, paragraf, atau dokumen, menjadi token atau bagian tertentu berdasarkan setiap kata yang mengikutinya.

Tabel 2. Contoh proses tokenisasi

No.	Data sebelum proses tokenisasi	Data sesudah proses tokenisasi
1.	barang yang datang beda sekali sama yang di gambar.. customer amat sangat kecewa..	“barang”, “yang”, “datang”, “beda”, “sekali”, “sama”, “yang”, “di”, “gambar”, “customer”, “amat”, “amat”, “sangat”, “kecewa”

- C. Filtering (Stopword) merupakan proses menghilangkan kata-kata yang sering muncul namun tidak mempengaruhi analisis sentimen. Contoh yang dikenali adalah “dan”, “di”, “dari”, “yang”, “ke” dan seterusnya.

Tabel 3. Contoh Proses Filtering

No.	Data sebelum proses Stopword	Data sesudah proses stopwords
1.	barang yang datang beda sekali sama yang di gambar.. customer amat sangat kecewa..	barang datang beda sekali sama gambar..customer kecewa..

- D. Stemming adalah proses menghilangkan seluruh kata yang ada pada kata tersebut menjadi kata dasar. Misalnya saja yaitu meng-, me-, kan-, di-, pe-, peng-, dan sebagainya. Stemming bertujuan untuk mereduksi variasi kata yang mempunyai kata dasar yang sama.

Tabel 4. Contoh Proses Stemming

No.	Stemming Data sebelum diproses	Stemming data setelah diproses
1.	Barang yang datang berbeda dengan barang yang sama di gambar. Sangat mengecewakan.	Barang datang beda sesuai gambar. Pelanggan kecewa.

Pada tahap review produk selanjutnya, data komentar di Microsoft Excel dimasukkan ke dalam proses rapid miner. Process Clustering menggunakan 2 cluster yang terdiri dari cluster 0 dan cluster 1 yang menjelaskan bahwa cluster 0 adalah komentar negatif dan cluster 1 adalah komentar positif. Hasil yang diperoleh pada proses clustering antara lain cluster 0 berjumlah 116 data komentar review produk, dan cluster 1 berjumlah 37 data komentar review produk

## 4. KESIMPULAN

Dari hasil penelitian yang telah dilakukan maka diperoleh kesimpulan, dengan algoritma Naïve Bayes dapat memprediksi kelas sentimen pada ulasan produk di toko Kinophonecell pada platform e-commerce Shopee. 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%. Hasil analisis penelitian ini menghasilkan sentimen positif lebih unggul berjumlah 196 dan jumlah sentimen negatif berjumlah 4. Jumlah data latih dalam sistem memiliki pengaruh terhadap prediksi sistem. Selain jumlah, kualitas data latih juga berperan

### **Saran**

Bagi pelanggan yang ingin membeli produk agar tidak berpaku pada rating produk saja, namun melihat review pelanggan yang telah membeli produk di Spatuafa. Kemudian dari toko Spatuafa untuk dijadikan referensi dalam meningkatkan kualitas produk yang akan dipasarkan di website e-commerce Shopee. Agar produk yang dibeli pelanggan dapat memuaskan pelanggan.

### **REFERENSI**

- [1] Darwis, D., Siskawati, N., & Abidin, Z., "Penerapan Algoritma Naive Bayes Untuk Analisis Sentimen Review Data Twitter Bmkg Nasional," J. Tekno Kompak, 2021.
- [2] Amrullah, A. Z., Sofyan Anas, A., and Hidayat, M. A. J., "Analisis Sentimen Movie Review Menggunakan Naive Bayes Classifier Dengan Seleksi Fitur Chi Square," BITE, 2020.
- [3] Asri Amaliza Fathia Mausea, I. A. S., "RANCANG BANGUN APLIKASI PENDAFTARAN PASIEN ONLINE DAN PEMERIKSAAN DOKTER DI KLINIK PENGOBATAN BERBASIS WEB," Rekayasa Informasi, 2021.
- [4] A.R. Isnain, A. Sihabuddin, and Y. Suyanto, "Bidirectional Long Short Term Memory Method and Word2vec Extraction Approach for Hate Speech Detection," IJCCS, 2020.
- [5] A. Surahman, A. F. O. P. Pasaribu, and D. Darwis, "Ekstraksi Data Produk e-Marketplace Sebagai Strategi Pengolahan Segmentasi Pasar Menggunakan Web Crawler," Sist. J. Sist. Inf., 2020.
- [6] A. Surahman, "PENGEMBANGAN MARKET SEGMENTASI UNTUK MENCAPAI KEUNGGULAN BERSAING PADA E-MARKETPLACE," J. Komput. dan Inform., 2020.
- [7] A. Rahman, E. Utami, and S. Sudarmawan, "Sentimen Analisis Terhadap Aplikasi pada Google Playstore Menggunakan Algoritma Naïve Bayes dan Algoritma Genetika," J. Komtika (Komputasi dan Inform., vol. 5, no. 1, pp. 60–71, 2021, doi: 10.31603/komtika.v5i1.5188.
- [8] M. Joshi, H. Vala. "Opinion Mining For Sentiment Data Classification". International Journal of Research in Information Technology, 3 (1), 1-13. 2014..
- [9] L. Agusta. "Comparison algorithm Stemming Porter with Algoritular Nazief & Adriani for Stemming Bahasa Indonesia text document". National Conference of Systems and Informatics. Kns & 109-036. 2009.
- [10] D. Alita, Y. Fernando, and H. Sulistiani, "Implementasi Algoritma Multiclass SVM pada Opini Publik Berbahasa Indonesia di Twitter," J. Tekno Kompak, 2020.