

SENTIMEN ANALISIS TERHADAP CUSTOMER REVIEW PRODUK SHOPEE BERBASIS WORDCLOUD DENGAN ALGORITMA NAÏVE BAYES CLASSIFIER

SENTIMENT ANALYSIS OF CUSTOMER REVIEWS OF SHOPEE PRODUCTS BASED ON WORDCLOUD USING NAÏVE BAYES CLASSIFIER ALGORITHM

Anggi Hanafiah¹, Arbi Haza Nasution², Yudhi Arta³, Rizky Wandri⁴, Hafiza Oktasia Nasution⁵, Jerika Mardafora⁶

^{1,2,3,4,5,6}Universitas Islam Riau
anggihanafiah@eng.uir.ac.id

ABSTRACT

Product reviews are information that is needed to gain customer trust in the marketplace. Business people must carry out an analysis of the reviews provided by customers on product and service evaluations. In this study the method used is sentiment analysis. Sentiment analysis is a field in Natural Language Processing that builds a system to recognize and extract opinions in text form both from various internet data sources and various social media platforms. Sentiment analysis can identify the strengths and weaknesses in a business by understanding and classifying sentiments (positive, negative, neutral) contained in writing using text analysis. In this study, sentiment analysis will be carried out using the Naïve Bayes algorithm by determining one of the products on the Shopee marketplace. Then, using the wordcloud method, you can find out the positive and negative sentiments that are often given by customers. Based on the results of the implementation and testing of sentiment analysis, it can be concluded that the application of sentiment analysis is very effective and efficient for knowing the results of sentiment towards products and services through reviews or comments provided by customers with a large number of datasets. The information obtained can be used as evaluation material in optimizing these products and services to increase customer confidence in business people.

Keywords : *Sentiment Analysis, Shopee, Naïve Bayes, Marketplace*

ABSTRAK

Review produk merupakan informasi yang sangat dibutuhkan untuk mendapatkan kepercayaan customer pada marketplace. Pelaku bisnis harus perlu melakukan analisis dari review yang diberikan oleh customer terhadap evaluasi produk dan layanan. Dalam penelitian ini metode yang digunakan menggunakan metode sentiment analysis. Sentiment analysis merupakan salah satu bidang dalam Natural Language Processing yang membangun sistem untuk mengenali dan mengestrak opini dalam bentuk teks baik dari berbagai sumber data internet dan beragam platform media sosial. Dengan sentiment analysis dapat diketahui apa saja kekurangan dan keunggulan dalam sebuah bisnis dengan memahami dan mengelompokkan sentimen (positif, negative, netral) yang terdapat dalam tulisan dengan menggunakan analisis teks. Pada penelitian ini akan dilakukan analisis sentiment dengan menggunakan algoritma naïve bayes dengan menentukan salah satu produk pada marketplace shopee. Kemudian selanjutnya dengan metode wordcloud dapat mengetahui sentimen positif dan negative yang sering diberikan oleh customer. Berdasarkan hasil implementasi dan pengujian sentiment analysis ini dapat disimpulkan bahwa penerapan analisis sentimen ini sangat efektif dan efisien untuk mengetahui hasil sentimen terhadap produk dan layanan melalui review atau komentar yang diberikan customer dengan jumlah dataset yang besar. Informasi yang didapatkan dapat menjadi bahan evaluasi dalam mengoptimalkan produk dan layanan tersebut guna meningkatkan kepercayaan customer terhadap pelaku bisnis.

Kata Kunci: *Sentiment Analysis, Shopee, Naïve Bayes, Marketplace*

PENDAHULUAN

Kegiatan jual beli barang dan jasa secara online atau yang biasa disebut dengan e-commerce telah berkembang pesat saat ini, terutama pada masa pandemic covid-19 (Purwati & Hamzah, 2022). E-commerce menawarkan pengguna

dalam melakukan sebuah transaksi barang maupun jasa tanpa batas jarak dan waktu selama pengguna memiliki akses internet (Işoraité & Miniotiené, 2018; Praniffa et al., 2023). Marketplace menjadi pilihan yang sangat digemari masyarakat dikarenakan banyaknya kelebihan-kelebihan antara lain

customer tidak perlu datang ke tempat, bisa diakses kapanpun, biaya yang relatif murah, diskon yang banyak, serta pembelian tanpa ada ongkos kirim. Dengan kelebihan tersebut, tidak sedikit permasalahan yang muncul seperti seperti barang yang dikirim tidak sesuai dengan pesanan, barang yang dikirimkan tidak sesuai dengan deskripsi, kemasan yang rusak, dan sebagainya. Hal ini yang menjadikan setiap customer ingin memberikan review terhadap kualitas produk dan layanan diberikan kepada pelaku bisnis. Sehingga dengan adanya review tersebut dapat mempengaruhi kepercayaan masyarakat atau customer untuk belanja secara online (Muktafin et al., 2020; Rohman et al., 2020; Nengsih et al., 2021).

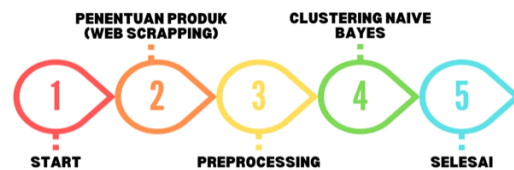
Tingkat kepercayaan customer terhadap penjual menjadi tantangan bagi para pelaku bisnis agar mampu melakukan evaluasi terhadap produk dan layanannya. Pelaku bisnis perlu mengidentifikasi masalah yang muncul dari review yang diberikan oleh para customer. Namun, membaca dan mengklasifikasikan setiap review yang diberikan oleh para customer membutuhkan waktu lama, dan dianggap tidak efektif (Sihombing et al., 2021). Penelitian yang dilakukan oleh Lutfi dan Permatasari menganalisis ulasan produk di marketplace Bukalapak untuk mengetahui sentimen ulasan pengguna bernilai positif atau negatif dengan pendekatan memanfaatkan Support Vector Machine dengan akurasi 93,42% (Lutfi et al., 2018).

Sentimet analysis ialah proses mengekstraksi, mengolah dan memahami data berupa teks yang tidak terstruktur secara otomatis guna mengambil informasi sentimen yang terdapat pada sebuah kalimat pendapat atau opini (Brahimi et al., 2021). Naïve Bayes merupakan salah satu algoritma yang mampu melakukan pengklasifikasian dengan cepat. Naïve Bayes juga merupakan salah satu algoritma yang sangat efisien dan efektif bahkan saat digunakan untuk menganalisis data berskala besar (Ahmad et al., 2017). Pada penelitian ini akan dilakukan sentiment

analysis dengan menggunakan algoritma naïve bayes dengan menentukan salah satu produk pada marketplace shopee. Kemudian selanjutnya dengan metode wordcloud dapat mengetahui sentiment positif dan negative yang sering diberikan oleh customer

METODE

Metode yang digunakan pada penelitian dalam analisis sentimen ini menggunakan pendekatan Natural Language Processing (NLP) dengan penerapan algoritma naïve bayes (Sidik et al., 2023). Tahapan yang dilakukan pada penelitian ini yaitu dengan melakukan penentuan produk di marketplace shopee, kemudian melakukan pre-processing terhadap produk yang sudah ditentukan, serta melakukan proses clustering untuk menentukan jumlah hasil dari klasifikasi sentiment. Adapun tahapan penelitian yang dilakukan dalam analisis sentimen dapat dilihat pada gambar 1 berikut.



Gambar 1. Tahapan Penelitian

Collecting Data

Pada proses pengumpulan data ini, dataset dilakukan dengan metode web scrapping dari website marketplace shopee menggunakan bahasa pemrograman python. Web scraper adalah cara untuk mendapatkan banyak data dari website dengan lebih cepat dan dalam waktu yang singkat (Ransi et al., 2020). Produk yang dijadikan sampel untuk penelitian ini yaitu produk smartphone dengan merek iphone xsmax pada bulan desember 2022. Setelah data berhasil di scrapping, data tersebut disimpan dalam format csv. Data yang berhasil dikumpulkan sebanyak 1754 dataset.

Preprocessing

Tahapan text preprocessing dibutuhkan untuk mengubah teks dari bentuk yang tidak terstruktur menjadi bentuk terstruktur yang akan digunakan kedalam proses text mining (Audrin & Audrin, 2022). Tahapan preprocessing ini terdiri dari beberapa tahapan yaitu :

1. Case Folding, yaitu tahapan yang dilakukan untuk mengubah semua huruf dalam dokumen menjadi huruf kecil (lowercase).
2. Cleaning, yaitu tahapan yang dilakukan untuk menghapus url, tanda baca, emoticon, angka, dan lainnya yang di anggap tidak penting.
3. Tokenizing, yaitu tahapan yang dilakukan untuk memecahkan gabungan kalimat menjadi per kata.
4. Stopword Removal, yaitu tahapan yang dilakukan untuk menghapus kata-kata yang tidak penting.
5. Steaming, yaitu tahapan yang dilakukan untuk merubah kata kedalam bentuk baku.

Term Weighting

Term Weighting atau pembobotan kata dilakukan setelah preprocessing data komentar akan dilakukan pembobotan kata untuk menghitung bobot dari setiap kata yang ada pada dokumen. Metode yang dapat digunakan salah satunya adalah TF-IDF (Term Frequency - Inverse Document Frequency). Bobot dari setiap data merupakan hasil perkalian dari TF dan IDF. TF merupakan frekuensi jumlah term pada data, dan IDF adalah fungsi untuk mengurangi bobot term jika kemunculannya banyak tersebar dalam dokumen.

$$Wt, d = tf_{(t,d)} \times idf_t$$

dengan :

$$idf_t = \log\left(\frac{N}{dft}\right) + 1$$

$$tf(t, d) = \frac{F_{t,d}}{\text{Number of word } d}$$

dengan :

- Ft,d = Frekuensi kemunculan term.
- N = Jumlah kata pada dokumen.

- idf t = Inverse Document Frecuency.
- dft = Total term pada dokumen.

Naïve Bayes Classifier

Proses klasifikasi pada penelitian ini menggunakan metode naïve bayes. Naive Bayes merupakan salah satu algoritma pembelajaran mesin probabilistik berdasarkan teorema bayes, yang biasa digunakan dalam berbagai macam tugas klasifikasi (Ratnawati, 2018). Secara umum naïve bayes memiliki teorema sebagai berikut :

$$P(Cj|Wi) = \frac{P(Wi|Cj)P(Cj)}{P(Wi)}$$

dengan :

- P(Cj|Wi) : Peluang kategori j, ketika terdapat kemunculan kata i.
- P(wi|cj) : Peluang kata i masuk kedalam kategori.
- P(Cj) : Peluang kemunculan kategori j.
- P(Wi) : Peluang kemunculan kata.
- Untuk menghitung P(Cj|Wi), dapat dilakukan dengan rumus berikut :

$$P(Cj|Wi) = \frac{Nic + 1}{Nc + V}$$

dengan :

- Nic : Jumlah dokumen latih dengan fitur atribut Wi dengan kategori Cj
- NC : Jumlah dokumen dari kategori Cj
- V = Jumlah kategori

HASIL DAN PEMBAHASAN

Data yang digunakan pada penelitian ini merupakan data komentar pelanggan yang di scrapping pada salah satu produk di website e-commerce shopee menggunakan bahasa pemrograman python, kemudian hasil scrapping tersebut kemudian disimpan dalam format csv. Jumlah data yang dihasilkan sebanyak 1754 records dan lalu dilakukan tahapan preprocessing hingga menjadi 850 records. Contoh data yang dihasilkan dari web scrapping tersebut adalah sebagai berikut :

Tabel 1. Hasil Web Scrapping

No	Username	Review
1	wahidkurn	Akhirnya bisa nilai jga, setelah pemakaian sebulan ini aman ² aja BH tetep di 100% dari awal datang sampe skrg. Aga

curiga sih tapi gpp bagus
aja cmn batrenya cepat
habis.. Selebih nya oke
Soal IMEI sayangnya
gada garansi jadi aga
khawatir soal IMEI..
Worth it sih ini buat
dibeli dengan harga
segitu👍

Setelah mendapatkan hasil scrapping tersebut, maka selanjutnya dilakukan proses preprocessing. Pada proses preprocessing, tahapan pertama yang dilakukan dengan yaitu tahapan case folding. Tahapan case folding yaitu mentransformasikan semua huruf pada review menjadi huruf kecil semua. Contoh hasil tahapan case folding dapat dilihat pada tabel 2 berikut.

Tabel 2. Tahapan Case Folding

akhirnya bisa nilai jga, setelah pemakaian sebulan ini aman² aja bh tetep di 100% dari awal datang sampe skrg. aga curiga sih tapi gpp bagus aja cmn batrenya cepat habis.. selebih nya oke soal imei sayangnya gada garansi jadi aga khawatir soal imei.. worth it sih ini buat dibeli dengan harga segitu👍

Setelah melakukan tahapan case folding, selanjutnya melakukan tahapan cleaning. Tahapan cleaning yaitu menghapus semua URL, tanda baca, emoticon, dan angka yang tidak digunakan untuk analisis. Contoh hasil tahapan cleaning dapat dilihat pada tabel 3 berikut.

Tabel 3. Tahapan Cleaning

akhirnya bisa nilai jga setelah pemakaian sebulan ini aman aja bh tetep di 100 dari awal datang sampe skrg aga curiga sih tapi gpp bagus aja cmn batrenya cepat habis selebih nya oke soal imei sayangnya gada garansi jadi aga khawatir soal imei worth it sih ini buat dibeli dengan harga segitu

Setelah data sudah bersih dari noise, maka selanjutnya melakukan tahapan tokenizing. Tahapan tokenizing yaitu memecah suatu kalimat menjadi token-token yang dipisahkan berdasarkan spasi. Contoh hasil tahapan tokenizing dapat dilihat pada tabel 4 berikut.

Tabel 4. Tahapan Tokenizing

['akhirnya', 'bisa', 'nilai', 'jga', 'setelah', 'pemakaian', 'sebulan', 'ini', 'aman', 'aja', 'bh', 'tetep', 'di', '100', 'dari', 'awal', 'datang', 'sampe', 'skrg', 'aga', 'curiga', 'sih', 'tapi', 'gpp', 'bagus', 'aja', 'cmn', 'batrenya', 'cepat', 'habis', 'selebih', 'nya', 'oke', 'soal', 'imei', 'sayangnya', 'gada', 'garansi', 'jadi', 'aga', 'khawatir', 'soal', 'imei', 'worth', 'it', 'sih', 'ini', 'buat', 'dibeli', 'dengan', 'harga', 'segitu', '']

Tahap yang dilakukan selanjutnya yaitu tahapan stopword removal. Tahapan ini dilakukan untuk menyeleksi kata-kata yang tidak memiliki kaitannya dengan analisis sentiment. Contoh hasil tahapan stopword removal dapat dilihat pada tabel 5 berikut.

Tabel 5. Tahapan Stopword Removal

['nilai', 'pemakaian', 'sebulan', 'aman', 'bh', 'tetep', '100', 'sampe', 'skrg', 'curiga', 'gpp', 'bagus', 'cmn', 'batrenya', 'cepat', 'habis', 'selebih', 'imei', 'sayangnya', 'garansi', 'khawatir', 'imei', 'worth', 'it', 'dibeli', 'harga', 'segitu', '']

Setelah melakukan penyeleksian kata, selanjutnya dilakukan tahapan stemming. Tahapan stemming dilakukan untuk merubah kata yang diseleksi sebelumnya menjadi sebuah kata baku. Contoh hasil tahapan stemming dapat dilihat pada tabel 6 berikut.

Tabel 6. Tahapan Stemming

['nilai', 'pakai', 'bulan', 'aman', 'bh', 'tetep', '100', 'sampe', 'skrg', 'curiga', 'gpp', 'bagus', 'cmn', 'batrenya', 'cepat', 'habis', 'lebih', 'imei', 'sayang', 'garansi', 'khawatir', 'imei', 'worth', 'it', 'beli', 'harga', 'segitu', '']

Untuk tahapan analisis data, maka data hasil preprocessing diatas diperlukan proses labelling data. Proses labelling data digunakan untuk mengetahui polaritas terhadap review pelanggan. Dari jumlah data yang dihasilkan pada proses preprocessing didapatkan data sebanyak 585 sentiment positif, 170 sentimen netral, dan 95 sentimen negatif. Hasil sentimen

pengujian menggunakan data training sebesar 70% mendapat nilai akurasi sebesar 0.84%, hasil pengujian menggunakan data training sebesar 80% mendapat nilai akurasi sebesar 0.83%, dan hasil pengujian menggunakan data training sebesar 90% mendapat nilai akurasi sebesar 0.81%.

Dari hasil pengujian yang dilakukan dapat dilihat bahwa pembagian data yang diuji menggunakan data training 60% menggunakan fitur pembobotan TF-IDF dengan library scikit learn memperoleh tingkat akurasi paling tinggi.

DAFTAR PUSTAKA

- Audrin, C., & Audrin, B. (2022). Key factors in digital literacy in learning and education: a systematic literature review using text mining. *Education and Information Technologies*, 27(6), 7395-7419.
- Ahmad, M., Aftab, S., Muhammad, S. S., & Ahmad, S. (2017). Machine Learning Techniques for Sentiment Analysis: A Review. *International Journal of Multidisciplinary Sciences and Engineering*, 8(3), 27–32. <http://www.ijmse.org/Volume8/Issue3/paper5.pdf>
- Brahimi, B., Touahria, M., & Tari, A. (2021). Improving sentiment analysis in Arabic: A combined approach. *Journal of King Saud University - Computer and Information Sciences*, 33(10), 1242–1250. <https://doi.org/10.1016/j.jksuci.2019.07.011>
- Išoraitė, M., & Miniotienė, N. (2018). Electronic Commerce: Theory and Practice. *IJBE (Integrated Journal of Business and Economics)*, 2(2), 73. <https://doi.org/10.33019/ijbe.v2i2.78>
- Lutfi, A. A., Permanasari, A. E., & Fauziati, S. (2018). Sentiment Analysis in the Sales Review of Indonesian Marketplace by Utilizing Support Vector Machine. *Journal of Information Systems Engineering and Business Intelligence*, 4(1), 57. <https://doi.org/10.20473/jisebi.4.1.57-64>
- Muktafin, E. H., Kusriani, K., & Luthfi, E. T. (2020). Analisis Sentimen pada Ulasan Pembelian Produk di Marketplace Shopee Menggunakan Pendekatan Natural Language Processing. *Jurnal Eksplora Informatika*, 10(1), 32–42. <https://doi.org/10.30864/eksplora.v10i1.390>
- Nengsih, W., Zein, M. M., & Hayati, N. (2021). Coarse-grained sentiment analysis berbasis natural language processing—ulasan hotel. *Jurnal Nasional Teknik Elektro dan Teknologi Informasi*, 10(1), 41-48.
- Praniffa, A. C., Syahri, A. ., Sandes, F., Fariha, U., Giansyah, Q. A., & Hamzah, M. (2023). Pengujian Sistem Informasi Parkir Berbasis Web Pada UIN SUSKA RIAU Menggunakan White Box dan Black Box Testing. *Jurnal Testing Dan Implementasi Sistem Informasi*, 1(1), 1-6. Retrieved from <https://journal.almatani.com/index.php/jtisi/article/view/321>
- Purwati, A. A., & Hamzah, M. L. (2022). Instrument Readiness Analysis of Technology-Based Businesses. *Journal of Applied Engineering and Technological Science (JAETS)*, 4(1), 611-617.
- Ransi, N., Surimi, L., Tenriawaru, A., & Saidi, L. O. (2020). Analisis Sentimen Masyarakat Terhadap Toko Online Aplikasi. 1–8.
- Ratnawati, F. (2018). Implementasi Algoritma Naive Bayes Terhadap Analisis Sentimen Opini Film Pada Twitter. *INOVTEK Polbeng-Seri Informatika*, 3(1), 50–59.
- Rohman, A. N., Musyarofah, R. L., Utami, E., & Raharjo, S. (2020, October). Natural Language Processing on Marketplace Product Review Sentiment Analysis. In *2020 2nd International Conference on Cybernetics and Intelligent System*

(*ICORIS*) (pp. 1-5). IEEE.

- Sidik, F., Suhada, I., Anwar, A. H., & Hasan, F. N. (2022). Analisis Sentimen Terhadap Pembelajaran Daring Dengan Algoritma Naive Bayes Classifier. *Jurnal Linguistik Komputasional*, 5(1), 34-43.
- Sihombing, L. O., Hannie, H., & Dermawan, B. A. (2021). Sentimen Analisis Customer Review Produk Shopee Indonesia Menggunakan Algoritma Naive Bayes Classifier. *Edumatic: Jurnal Pendidikan Informatika*, 5(2), 233-242. <https://doi.org/10.29408/edumatic.v5i2.4089>