



## Analisis Sentimen Ulasan Produk di *E-Commerce* Bukalapak Menggunakan *Natural Language Processing*

### *Sentiment Analysis of Product Reviews on E-Commerce's Bukalapak Using Natural Language Processing*

Elsa Sera<sup>1</sup>, Hazriani<sup>2</sup>, Mirfan<sup>3</sup>, Yuyun<sup>4</sup>

<sup>1,2</sup>Sistem Komputer, Program Pasca Sarjana, Universitas Handayani Makassar

<sup>3</sup>Teknik Informatika, Ilmu Komputer, Universitas Handayani Makassar

<sup>4</sup>Badan Riset dan Inovasi (BRIN)

<sup>1</sup>elsasera07@gmail.com

#### **Abstract**

*This research discusses the analysis of sentiment in product reviews on E-Commerce Bukalapak using Natural Language Processing (NLP). The study aims to fill the knowledge gap regarding the analysis of product reviews in online stores in Indonesia, specifically Bukalapak. The data used in this research were collected from various product categories, such as clothing, electronics, cosmetics, and others. The method employed in this study was the TF-IDF method to train the Naive Bayes model. The results of the research show that the Naive Bayes model trained using the TF-IDF method achieved an accuracy of 88%. This indicates that the model has good capability in predicting the sentiment of product reviews. The analysis of positive reviews reveals customer satisfaction with product quality, fast delivery, reasonable pricing, and receiving items as expected. On the other hand, the analysis of negative reviews uncovers the mismatch between customer expectations and the actual conditions regarding color, delivery, and product orders. This study contributes to a deeper understanding of sentiment analysis in product reviews on E-Commerce Bukalapak. The insights from this analysis can be utilized by Bukalapak to enhance the quality of their products and services, providing a more satisfying experience for customers.*

*Keywords: sentiment analysis, product reviews, E-Commerce Bukalapak, Natural Language Processing, Naive Bayes model, TF-IDF method.*

#### **Abstrak**

penelitian ini membahas tentang analisis sentimen ulasan produk di *E-Commerce* Bukalapak menggunakan *Natural Language Processing* (NLP). Penelitian ini bertujuan untuk mengisi celah pengetahuan terkait analisis ulasan produk di toko online di Indonesia, khususnya Bukalapak. Data yang digunakan dalam penelitian ini diambil dari berbagai kategori produk, seperti pakaian, elektronik, kosmetik, dan lain sebagainya. Metode yang digunakan dalam penelitian ini adalah metode TF-IDF untuk melatih model *Naive Bayes*. Hasil penelitian menunjukkan bahwa model *Naive Bayes* yang dilatih menggunakan metode TF-IDF memiliki akurasi sebesar 88%. Hal ini menunjukkan bahwa model memiliki kemampuan yang baik dalam memprediksi sentimen ulasan produk. Analisis ulasan positif menunjukkan adanya kepuasan pembeli terhadap kualitas produk, kecepatan pengiriman, harga yang baik, dan barang yang diterima sesuai harapan. Sementara itu, analisis ulasan negatif mengungkapkan adanya ketidakcocokan antara harapan pembeli dengan kondisi sebenarnya terkait warna, pengiriman, dan pesanan produk. Penelitian ini memberikan kontribusi dalam pemahaman lebih lanjut mengenai analisis sentimen ulasan produk di *E-Commerce* Bukalapak. Informasi hasil analisis ini dapat digunakan oleh Bukalapak untuk meningkatkan kualitas produk dan layanan mereka, serta memberikan pengalaman yang lebih memuaskan bagi pelanggan.

Kata kunci: Analisis sentimen, ulasan produk, *E-Commerce* Bukalapak, *Natural Language Processing*, model *Naive Bayes*, metode TF-IDF.

#### **1. Pendahuluan**

Dalam era digital saat ini, toko online telah menjadi salah satu pilihan utama konsumen dalam membeli produk.[1] Namun, dengan banyaknya produk yang tersedia di toko online dan minimnya interaksi langsung dengan produk, ulasan produk dari konsumen menjadi penting sebagai sumber informasi tambahan untuk

membuat keputusan pembelian. Oleh karena itu, analisis sentimen ulasan produk menjadi relevan dalam membantu konsumen memilih produk yang tepat.[2]

Menurut statistik terbaru mengenai ulasan online, hampir setengah (49 persen) dari semua konsumen mengungkapkan bahwa mereka memiliki tingkat kepercayaan yang sama terhadap ulasan online seperti

rekomendasi pribadi. Data dari BrightLocal (2022) menunjukkan bahwa ulasan online menjadi sumber informasi yang signifikan bagi konsumen dalam membuat keputusan pembelian. Dalam survei yang dilakukan, tingkat kepercayaan konsumen terhadap ulasan online juga dibandingkan dengan sumber informasi lain. Hasilnya menunjukkan bahwa 37 persen konsumen memiliki tingkat kepercayaan yang setara terhadap ulasan online dan rekomendasi yang mereka lihat di media sosial. Di sisi lain, 28 persen konsumen menyatakan bahwa mereka mempercayai ulasan online sebanyak artikel yang ditulis oleh pakar subjek. Data ini menunjukkan bahwa ulasan online memiliki pengaruh yang signifikan dalam membentuk kepercayaan konsumen. Konsumen cenderung mengandalkan ulasan online sebagai sumber informasi yang dapat diandalkan dalam membuat keputusan pembelian, bahkan sejajar dengan rekomendasi pribadi dan konten media sosial.[3]

Penelitian terkait analisis sentimen ulasan produk telah dilakukan sebelumnya oleh banyak peneliti.[4] Beberapa penelitian tersebut telah menggunakan teknik *Natural Language Processing* (NLP) untuk melakukan analisis sentimen pada ulasan produk.[5] Beberapa penelitian juga telah memfokuskan pada toko online tertentu, seperti Amazon, eBay, dan lain-lain. Namun, penelitian yang fokus pada toko online di Indonesia masih terbatas.[6] Oleh karena itu, Penelitian ini dilakukan untuk mengisi celah pengetahuan terkait analisis ulasan produk toko online di Indonesia, dengan fokus pada *E-Commerce* Bukalapak. Data diambil dari berbagai kategori produk, seperti pakaian, elektronik, kosmetik, dan lain sebagainya. Data yang berhasil diambil kemudian dilakukan preprocessing untuk mempersiapkan data sebelum dilakukan analisis sentimen dengan menggunakan *Natural Language Processing*. [7]

Dengan demikian, hasil penelitian ini dapat memberikan kontribusi pada pengembangan model analisis sentimen ulasan produk yang dapat digunakan pada toko online di Indonesia. Dalam konteks bisnis, hasil penelitian ini dapat membantu toko online dalam meningkatkan kualitas produk dan layanan mereka dengan memahami sentimen dan kebutuhan pelanggan. Selain itu, konsumen juga dapat diuntungkan dengan adanya informasi tambahan yang lebih akurat dan bermanfaat dalam memilih produk.[8]

## 2. Metode Penelitian

Metode penelitian yang digunakan dalam studi ini adalah metode analisis sentimen menggunakan *Natural Language Processing* (NLP) [9] yang terdiri dari beberapa tahapan seperti yang ditunjukkan pada Gambar 1. Tahapan awal penelitian ini dimulai dengan penentuan produk yang akan dianalisis, lalu dilanjutkan dengan pengumpulan data ulasan produk. Dilakukan pra-pemrosesan pada ulasan yang sudah terkumpul dengan dan tanpa pendekatan NLP sehingga menghasilkan 2 dataset, dengan pendekatan NLP dan

dataset tanpa pendekatan NLP. Setiap dataset dibagi menjadi data latih dan data uji, kemudian diterapkan perhitungan pembobotan TF-IDF pada kedua dataset untuk dibandingkan hasil keduanya dan dilakukan analisis hasil.



Gambar 1. Alur Penelitian

### 2.1. *Natural Language Processing* (NLP)

*Natural Language Processing* (NLP) adalah cabang ilmu komputer yang berkaitan dengan pemahaman dan pemrosesan bahasa manusia secara alami oleh komputer. NLP berfokus pada pengembangan metode dan algoritma untuk mengenali, memahami, menganalisis, dan memanipulasi teks dan bahasa manusia dalam bentuk yang dapat dipahami oleh mesin. Tujuan utama NLP adalah memungkinkan komputer untuk berinteraksi dengan manusia melalui bahasa manusia dengan cara yang lebih efektif dan intuitif.[10]

### 2.2. Sumber Data

Sumber data yang digunakan dalam studi ini adalah ulasan produk dari toko online Bukalapak yang diambil dari <https://www.kaggle.com/>. Dataset ini terdiri dari puluhan ribu ulasan produk yang terdiri dari dua kategori sentimen yaitu positif dan negatif. Dataset ini merupakan salah satu dari dataset terbesar dalam bahasa Indonesia yang digunakan untuk analisis sentimen ulasan produk menggunakan NLP.

### 2.3. Pra-pemrosesan

Dataset kemudian melalui proses pra-pemrosesan untuk membersihkan dan menyiapkan data sehingga siap dianalisis. Pra-pemrosesan dilakukan dalam 2 keadaan, yang pertama pra-pemrosesan tanpa pendekatan NLP yang terdiri dari tahapan menghilangkan emoticons dan simbol, *lowercase folding*, dan penanganan *tag*. Yang kedua pra-pemrosesan dengan pendekatan *Natural Language Processing* (NLP) dengan tujuan untuk memperbaiki bahasa yang ada pada ulasan, karena pada ulasan ditemukan banyak penggunaan kata yang tidak baku dan singkatan. *Natural Language Processing* (NLP) digunakan untuk mengatasi hambatan pengenalan bahasa percakapan sehari-hari oleh sistem komputer.[11]



Gambar 2. Tahap pra-pemrosesan dengan pendekatan NLP

Alur kerja tahapan pra-pemrosesan dengan pendekatan NLP ditunjukkan pada Gambar 2, dengan a.) Menghilangkan *emoticons* dan simbol yang ada pada ulasan dibuang karena penelitian ini hanya berfokus pada teks yang terkandung dalam ulasan. Simbol yang dibuang adalah “~”, “’”, “!”, “\$”, “%”, “^”, “&”, “\*”, “(”, “)”, “\_”, “+”, “=”, “:”, “;”, “<”, “>”, “koma”, “titik”, “?”, “/”, “\”, “#”, dan “|”. Ulasan yang mengandung simbol seperti “memuaskan banget kak : )” setelah diproses maka menjadi “memuaskan banget kak”. b.) *Lowercase folding* adalah mengubah semua huruf menjadi huruf kecil agar terhitung sebagai kata yang sama. Komentar “Bagus banget sesuai pesanan” menjadi “bagus banget sesuai pesanan”, huruf “B” kapital diubah menjadi “b” huruf kecil. c.) Penanganan *tag* ulasan adalah potongan informasi yang telah disediakan Bukalapak dan dapat disertakan pada ulasan pembelian produk. Tidak semua ulasan mengandung tag, sehingga perlu dilakukan penanganan khusus pada ulasan yang mempunyai tag. Tag pada ulasan ditambahkan menjadi bagian dari isi komentar ulasan. [12] Contoh tag yang disediakan bukalapak seperti “Kecepatan pengiriman sangat baik”, “Respon penjual sangat baik”. d.) *Word Normalizer* digunakan untuk memperbaiki kata-kata dalam ulasan sehingga menghasilkan kalimat yang baik dan benar sesuai dengan aturan tata bahasa Indonesia. Peningkatan ini diperlukan untuk memudahkan pembaca memahami makna kalimat. Komentar “barangnya bagus bgt” maka setelah proses *Word Normalizer* menjadi “barangnya bagus banget”. Kata “bgt” diubah menjadi “banget” sehingga lebih mudah dipahami. e.) *Stemming* berfungsi untuk membuat suatu kata menjadi kata dasar, dengan menghilangkan semua imbuhan yang ada pada kata tersebut. Sebagai contoh kalimat “bahannya halus dan tidak membuat kegerahan” kemudian diubah menjadi “bahan halus dan tidak buat gerah”. f.) *Stopword Removal* berfungsi untuk menghilangkan kata-kata yang mempunyai jumlah kemunculan banyak tapi tidak terlalu penting. Kata yang masuk *stopword* seperti “yang”, “dan”, “di”, “dari” sehingga menyisakan kata-kata yang penting. Sebagai contoh kalimat “barang yang warna hijau dan biru bagus” kemudian diubah menjadi “barang warna hijau biru bagus”. g.) *Tokenizer* berfungsi untuk membagi teks input menjadi array token, karakter yang digunakan dalam bahasa Indonesia adalah alfabet dan setiap token dipisahkan oleh spasi. Sebagai contoh kalimat “barangnya bagus banget” kemudian diubah menjadi “barangnya”, “bagus”, “banget”. [13]

#### 2.4. Term Frequency Inverse Document Frequency (TF-IDF)

Dataset yang telah melalui tahapan pra-pemrosesan dengan dan tanpa pendekatan NLP, dibagi menjadi 2 bagian yaitu data latih dan data uji. Kemudian diterapkan pembobotan dengan algoritma Term Frequency Inverse Document Frequency (TF-IDF). TF-IDF digunakan untuk memberikan pembobotan hubungan suatu kata atau istilah terhadap keseluruhan ulasan. Frekuensi kemunculan kata di dalam ulasan menunjukkan seberapa penting kata itu di dalam ulasan tersebut, dan ulasan mana yang memiliki kata tersebut sehingga ulasan dapat diklasifikasikan ke dalam 2 kelas (ulasan positif dan ulasan negatif). Perhitungan TF-IDF menggunakan Persamaan 1.

$$W_{x,y} = tf_{x,y} \times \log \frac{N}{df_x} \quad (1)$$

Dengan  $W_x$ , adalah bobot istilah (ty) terhadap dokumen (dx). Sedangkan  $tf_x$ , adalah jumlah kemunculan istilah (ty) dalam dokumen (dx). N adalah jumlah semua dokumen yang ada dalam dataset dan  $df_x$ , adalah jumlah dokumen yang mengandung istilah (ty), minimal ada satu kata yaitu istilah (ty). [11]

#### 2.5 Analisis Ulasan dari dataset

Dalam penelitian ini, dataset ulasan produk BukaLapak yang diperoleh dari Kaggle.com diolah menggunakan *Google Colab* dan bahasa pemrograman Python. Pertama, dataset ulasan produk BukaLapak yang telah diunduh dari Kaggle.com diunggah ke *Google Colab*. Hal ini memungkinkan peneliti untuk mengakses dan memanipulasi dataset dengan mudah menggunakan lingkungan pengembangan yang disediakan oleh *Google Colab*.

*Google Colab* memberikan lingkungan yang efisien dan fleksibel untuk melakukan analisis ulasan. Beberapa langkah yang dilakukan dalam analisis ulasan menggunakan *Google Colab* dan *Python* meliputi: Persiapan Data: Data ulasan yang digunakan dalam penelitian ini diunggah ke *Google Colab*. Data dapat berupa file CSV, Excel, atau format lainnya. Dalam penelitian ini, data tersebut diimpor menggunakan library *Pandas* untuk membaca dan memanipulasi dataset. Pemrosesan Teks: Setelah data terbaca, langkah selanjutnya adalah melakukan pemrosesan teks pada ulasan. Hal ini melibatkan tahap seperti tokenisasi (membagi teks menjadi kata-kata), normalisasi (misalnya, mengubah huruf menjadi huruf kecil), penghapusan kata-kata umum (*stop words*), dan pembersihan teks lainnya. Analisis Sentimen: Salah satu jenis analisis yang umum dilakukan adalah analisis sentimen, yang bertujuan untuk menentukan sentimen atau perasaan yang terkandung dalam setiap ulasan. Dalam penelitian ini, dapat digunakan *library* seperti *NLTK* atau *spaCy* untuk melakukan analisis sentimen. Analisis Frekuensi Kata: Analisis frekuensi kata melibatkan menghitung berapa kali kata-kata tertentu muncul dalam dataset ulasan. Hal ini dapat memberikan

wawasan tentang kata-kata yang paling umum digunakan oleh pengguna. Library seperti NLTK dapat digunakan untuk melakukan analisis frekuensi kata. Visualisasi Data: Setelah analisis dilakukan, langkah selanjutnya adalah memvisualisasikan hasilnya agar mudah dipahami. Dalam penelitian ini, library seperti Matplotlib atau Seaborn dapat digunakan untuk membuat plot atau grafik yang menjelaskan hasil analisis. Interpretasi Hasil: Setelah mendapatkan hasil analisis, penting untuk menginterpretasikan temuan dan menyajikannya secara jelas dan komprehensif.

### 3. Hasil dan Pembahasan

#### 3.1. Daftar Ulasan Produk Bukalapak

Berikut sampel dataset yang di ambil secara acak :

Tabel 1. contoh ulasan produk

No	id	tag	review	label
1	2	Barang Sesuai	Terima kasih buka lapaklapak barang sesuai dengan keinginan dan memuaskan	positif
2	40	Kondisi Barang	Barang sesuai foto. Tidak mengecewakan. Pengiriman tepat waktu. Okey semua	positif
3	129	Respon Lama Sekali ,Kecewa	Respon lama sekali .kecewa sdh bayar dua hari chat tdk ada balasan sama sekali	Negatif
4	101	Gak Sesuai Deskripsi	Apa yg ditulis bahwa ini grade ori dgn detail nya , ternyata ga sesuai ... Terimakasih ..	Negatif
5	30	Nice	Barang sudah sampai dgn selamat. thanks bl	positif

Berdasarkan Tabel 1. analisis yang di lakukan terhadap tabel di atas, dapat disimpulkan bahwa: Teks 1, 2, dan 5 menunjukkan sentimen positif, dengan penggunaan kata-kata seperti "barang sesuai foto," "pengiriman tepat waktu," dan "terima kasih." Hal ini mengindikasikan kepuasan terhadap produk dan pengiriman yang baik. Teks 3 menunjukkan sentimen negatif, karena terdapat ketidaksesuaian antara apa yang dijanjikan dalam deskripsi produk dengan apa yang diterima. Hal ini menyebabkan ketidakpuasan dan diungkapkan dalam frase "tidak sesuai" dan "terima kasih." Meskipun teks 4 menggunakan frase "terima kasih," namun terdapat ketidakpuasan yang diungkapkan melalui kalimat "tidak sesuai dengan yang ditulis." Oleh karena itu, teks ini dikategorikan sebagai sentimen negatif. Kami melakukan analisis ini berdasarkan penggunaan kata-kata dan frase yang mengindikasikan kepuasan atau ketidakpuasan, serta memperhatikan konteks yang terkandung dalam teks-teks tersebut.

#### 3.2. Analisis Ulasan

Ulasan yang digunakan sebanyak 200 ulasan, yang telah diberi label dan menghasilkan 178 ulasan positif dan 22 ulasan negatif. Ulasan dijadikan dataset dengan

komposisi 80:20 80% data sebagai data latih (160 dataset) dan 20% data sebagai data uji (40 dataset), kemudian dilakukan pra-pemrosesan dengan pendekatan NLP. Pada tahap pra-pemrosesan diterapkan fitur *word normalizer*, *stemming*, dan *stopword removal* untuk setiap ulasan. *Word normalizer* untuk menangani variasi penulisan kata yang memiliki makna yang sama agar terhitung sebagai istilah tunggal . Contoh variasi penulisan kata yang memiliki makna sama ditunjukkan pada Tabel 2.

Tabel 2. Variasi penulisan kata pada ulasan yang memiliki makna sama.

No.	Istilah pada ulasan	Word normalizer	Jumlah ulasan
1	tdk	tidak	6
2	gak	tidak	17
3	bgt	banget	3
4	Brng	barang	2
5	dtng	datang	2

Pada Tabel 2 menunjukkan variasi kata yang memiliki makna sama. Istilah-istilah tersebut termasuk "tdk" yang merupakan singkatan dari "tidak" dengan jumlah ulasan sebanyak 6, "gak" yang juga merupakan singkatan dari "tidak" dengan jumlah ulasan sebanyak 17, "bgt" yang merupakan singkatan dari "banget" dengan jumlah ulasan sebanyak 3, "Brng" yang merupakan singkatan dari "barang" dengan jumlah ulasan sebanyak 2, dan "dtng" yang merupakan singkatan dari "datang" dengan jumlah ulasan sebanyak 2. Penggunaan *word normalizer* dapat menangani banyaknya variasi kata yang digunakan pada penulisan ulasan oleh pembeli untuk diubah menjadi istilah yang sama.

Selanjutnya dilakukan proses *stemming* pada dataset, untuk menghilangkan awalan, sisipan dan akhiran kata sehingga menjadi bentuk dasarnya, dengan tujuan pengambilan informasi menjadi efisien dan efektif . Contoh penerapan stemming ditunjukkan pada Tabel 3.

Tabel 3. Hasil penerapan stemming

No	Istilah pada ulasan	stemming	Jumlah ulasan
1	dikirim	kirim	2
2	berlangganan	langgan	20
3	dijahit	jahit	25
4	pesanan	pesan	30
5	dipesan	pesan	32

Tabel 3. menunjukkan hasil penerapan stemming pada istilah-istilah yang ditemukan dalam ulasan, stemming yang dihasilkan, dan jumlah ulasan yang mengandung kata tersebut. Istilah-istilah tersebut termasuk "dikirim" yang diubah menjadi " kirim" dengan jumlah ulasan sebanyak 2, "berlangganan" yang diubah menjadi " langgan" dengan jumlah ulasan sebanyak 20, "dijahit" yang diubah menjadi " jahit" dengan jumlah ulasan sebanyak 25, "pesanan" yang diubah menjadi " pesan" dengan jumlah ulasan sebanyak 30, dan "dipesan" yang diubah menjadi " pesan" dengan jumlah ulasan sebanyak 32.

Selanjutnya dilakukan proses stopword removal untuk menghilangkan stopwords dari ulasan. Daftar stopwords yang digunakan dibuat sendiri mengacu pada konteks kata yang sering dipakai dalam ulasan maupun dalam jual beli secara daring. Contoh daftar stopwords yang dipakai ditunjukkan pada Tabel 4

Tabel 4. Contoh daftar stopwords

No	stopword	Jumlah kemunculan
1	yang	11
2	di	70
3	tapi	21
4	dan	28
5	untuk	6

Tabel 4 menampilkan stopwords, jumlah kemunculan *stopword* tersebut dalam dataset, yang mencakup kata "yang" dengan jumlah kemunculan sebanyak 11, "di" dengan jumlah kemunculan sebanyak 70, "tapi" dengan jumlah kemunculan sebanyak 21, "dan" dengan jumlah kemunculan sebanyak 28, dan "untuk" dengan jumlah kemunculan sebanyak 6. Stopword ini akan dihapus saat proses preprocessing untuk memperoleh data yang lebih bersih dan relevan dalam analisis sentimen.

Hasil keseluruhan proses pra-pemrosesan dapat dilihat pada Tabel 5.

Tabel 5. Hasil pra-pemrosesan

No	Ulasan Produk	Pra-pemrosesan				label aktual
		Tokenisasi	Normalisasi	topword Removal	Stemming / Lemmatization	
1	Terima kasih buka lapaklapak barang sesuai dengan keinginan dan memuaskan	"Terima", "kasih", "bukalapak", " ", "barang", "sesuai", "dengan", "keinginan", "dan", "memuaskan"	"terima", "kasih", "bukalapak", " ", "barang", "sesuai", "dengan", "keinginan", "dan", "memuaskan"	"terima", "kasih", " ", "barang", "keinginan", "memuaskan"	"terima", "kasih", " ", "barang", "keingin", "puas"	positif
2	Apa yg ditulis bahwa ini grade ori dgn detail nya , ternyata ga sesuai ... Terimakasih ,,	"Apa", "yang", "ditulis", "bahwa", "ini", "grade", "ori", "dengan", "detailnya", " ", "ternyata", "tidak", "sesuai", "..." "Terimakasih", " ", "	"apa", "yang", "tuliskan", "bahwa", "ini", "grade", "ori", "dengan", "detailnya", " ", "ternyata", "tidak", "sesuai", "..." "terimakasih", " ", "	"apa", "tuliskan", "grade", "ori", "detail", " ", "ternyata", "tidak", "sesuai", "..." kasih"	"apa", "tuliskan", "grade", "ori", "detail", " ", "ternyata", "tidak", "sesuai", "..." "terima", "kasih"	Negatif

Dimana pada Tabel 5 menunjukkan hasil pra-pemrosesan data ulasan produk. Pada tahap pra-pemrosesan, dilakukan beberapa teknik seperti normalisasi kata, stemming, dan penghapusan stopwords. Setiap baris dalam tabel mewakili satu ulasan produk beserta hasil pra-pemrosesan yang telah dilakukan. Kolom "Ulasan Produk" menampilkan teks asli dari ulasan produk yang diberikan oleh pengguna. Kolom "Pra-pemrosesan Word Normalizer" menunjukkan hasil normalisasi kata pada ulasan, di mana kata-kata seperti "buka lapaklapak" menjadi "buka lapak lapak" untuk memperbaiki kesalahan penulisan. Kolom "Pra-pemrosesan Stemming" menunjukkan hasil stemming pada ulasan, di mana kata-kata seperti "memuaskan" menjadi "muaskan" untuk mengambil akar kata. Kolom "Pra-pemrosesan Stopword Removal" menunjukkan hasil penghapusan stopwords pada ulasan, di mana kata-kata seperti "sesuai dengan" dan "dan" dihapus karena sifat umumnya. Kolom "Label" menunjukkan label sentimen yang diberikan untuk setiap ulasan setelah proses pra-pemrosesan. Terdapat label "positif" untuk ulasan yang menggambarkan kepuasan atau kesenangan, dan label "negatif" untuk ulasan yang

menggambarkan ketidakpuasan atau ketidaksesuaian. Dengan melakukan pra-pemrosesan pada ulasan produk, data telah disiapkan untuk analisis sentimen lebih lanjut dengan menggunakan teknik pembelajaran mesin.

### 3.3. Perhitungan model *Naive Bayes* Menggunakan TF-IDF

```

Akurasi model: 0.8780487804878049
Classification Report:
              precision    recall  f1-score   support

     0       0.00         0.00         0.00         5
     1       0.88         1.00         0.94        36

 accuracy         0.88         0.88         0.88         41
 macro avg       0.44         0.50         0.47         41
 weighted avg    0.77         0.88         0.82         41

Confusion Matrix:
[[ 0  5]
 [36  0]]

```

Gambar 3. Perhitungan Naive Bayes

Pada Gambar 3. dari hasil yang ditampilkan, dapat dilihat bahwa model *Naive Bayes* yang dilatih memiliki akurasi sekitar 0.88 atau sekitar 88%. Namun, perlu dicatat bahwa terdapat masalah dalam hasil prediksi karena model tidak dapat memprediksi label "0" (negatif) dengan baik, yang mengakibatkan precision, recall, dan f1-score untuk kelas "0" adalah 0. Model



Dalam kesimpulannya, penelitian ini menghasilkan model *Naive Bayes* dengan akurasi sekitar 88% dalam memprediksi sentimen ulasan produk di *E-Commerce* Bukalapak. Analisis ulasan positif menunjukkan bahwa pembeli merasa puas dengan kualitas, kecepatan pengiriman, dan harga produk yang diterima. Di sisi lain, analisis ulasan negatif mengungkapkan ketidaksesuaian antara harapan pembeli dan kondisi sebenarnya terkait warna, pengiriman, dan pesanan produk. Informasi ini dapat berguna bagi *E-Commerce* Bukalapak untuk meningkatkan kualitas produk dan layanan mereka, serta memberikan pengalaman yang lebih memuaskan bagi pelanggan.

2021, doi: 10.58496/mjbd/2021/002.

- [13] Warnia Nengsih, M. Mahrus Zein, and Nazifa Hayati, "Coarse-Grained Sentiment Analysis Berbasis *Natural Language Processing* – Ulasan Hotel," *J. Nas. Tek. Elektro dan Teknol. Inf.*, vol. 10, no. 1, pp. 41–48, 2021, doi: 10.22146/jnteti.v10i1.548.

## Daftar Rujukan

- [1] A. Hanafiah *et al.*, "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," vol. 6, pp. 230–236, 2023.
- [2] R. Siringoringo and J. Jamaludin, "Text Mining dan Klasterisasi Sentimen Pada Ulasan Produk Toko Online," *J. Teknol. dan Ilmu Komput. Prima*, vol. 2, no. 1, pp. 41–48, 2019, doi: 10.34012/jutikomp.v2i1.456.
- [3] B. Kumar Jha, "Sentiment Analysis for *E-Commerce* Products Using *Natural Language Processing*," *Ann. R.S.C.B.*, vol. 25, no. 5, pp. 166–175, 2021, [Online]. Available: <http://annalsofrscb.ro>
- [4] P. Muslim and A. C. Februadi, "Memprediksi Efek Electronic Word Of Mouth Terhadap Brand Image, Trust dan Behavioural Intention Pada Pelanggan Kedai Kopi Rumah Sekara," *Pros. Ind. Res. ....*, pp. 13–14, 2022, [Online]. Available: <https://jurnal.polban.ac.id/ojs-3.1.2/proceeding/article/view/4359%0Ahttps://jurnal.polban.ac.id/ojs-3.1.2/proceeding/article/download/4359/2873>
- [5] P. Arsi, R. Wahyudi, and R. Waluyo, "Optimasi SVM Berbasis PSO pada Analisis Sentimen Wacana Pindah Ibu Kota Indonesia," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 5, no. 2, pp. 231–237, 2021, doi: 10.29207/resti.v5i2.2698.
- [6] R. L. Atimi and Enda Esyudha Pratama, "Implementasi Model Klasifikasi Sentimen Pada Review Produk Lazada Indonesia," *J. Sains dan Inform.*, vol. 8, no. 1, pp. 88–96, 2022, doi: 10.34128/jsi.v8i1.419.
- [7] N. P. JATI, "S Integrasi Kansei Engineering Dan Kano Berbasis *Natural Language Processing* (Nlp) Untuk Menunjang Pengembangan Produk Jasa Pada Pariwisata Candi ...," 2021, [Online]. Available: <https://dspace.uui.ac.id/handle/123456789/37978%0Ahttps://dspace.uui.ac.id/bitstream/handle/123456789/37978/17522119.pdf?sequence=1>
- [8] H. Juliana, F. Ekonomi, and U. E. Unggul, "Pengaruh Siaran Langsung Dan Manfaat Yang Dirasakan Terhadap Niat Beli Melalui Kepercayaan Konsumen Pada Aplikasi Tiktok Shop," vol. 1, no. 6, pp. 1517–1538, 2023.
- [9] M. T. Anwar, D. Riandhita, A. Permana, P. Sistem, I. Industri, and J. Pusat, "Analisis Sentimen Masyarakat Indonesia Terhadap Produk Kendaraan Listrik Menggunakan VADER," vol. 10, no. 1, pp. 783–792, 2023.
- [10] M. Amien, "Sejarah dan Perkembangan Teknik *Natural Language Processing* (NLP) Bahasa Indonesia: Tinjauan tentang sejarah, perkembangan teknologi, dan aplikasi NLP dalam bahasa Indonesia," no. 2007, pp. 1–7, 2023, [Online]. Available: <http://arxiv.org/abs/2304.02746>
- [11] E. H. Muktafin, K. Kusrini, and E. T. Luthfi, "Analisis Sentimen pada Ulasan Pembelian Produk di Marketplace Shopee Menggunakan Pendekatan *Natural Language Processing*," *J. Eksplora Inform.*, vol. 10, no. 1, pp. 32–42, 2020, doi: 10.30864/eksplora.v10i1.390.
- [12] R. A. Hasan and T. Sutikno, "A Review on Big Data Sentiment Analysis Techniques," *Mesopotamian J. Big Data*, pp. 6–13,