

# Analisis Sentimen Produk Pada Bukalapak Menggunakan Lexicon-Based dan Multinomial Naïve Bayes

Brian Adam Bhagaskara<sup>a1</sup>, Ade Ismail<sup>a2</sup>, Habibie Ed Dien<sup>a3</sup>

<sup>a</sup> Program Studi Teknik Informatika, Jurusan Teknologi Informasi, Politeknik Negeri Malang  
Malang, Indonesia

<sup>1</sup>[brianadambhagaskara@gmail.com](mailto:brianadambhagaskara@gmail.com)

<sup>2</sup>[aismail@polinema.ac.id](mailto:aismail@polinema.ac.id)

<sup>3</sup>[habibie@polinema.ac.id](mailto:habibie@polinema.ac.id)

## Abstrak

Marketplace berperan penting dalam *e-commerce*, di mana reputasi produk sangat dipengaruhi oleh ulasan pembeli. Penelitian ini menganalisis sentimen ulasan "kemeja pria" di Bukalapak menggunakan *Natural Language Processing* (NLP). Data diproses melalui pre-processing seperti *cleaning*, *case folding*, normalisasi, *tokenizing*, *stopword removal*, dan *stemming*. Model *Lexicon-Based* dengan *InSet Lexicon* dan *Multinomial Naïve Bayes* menghasilkan *precision* 94%, *recall* 99% untuk sentimen positif, dan *accuracy* 91%-94%, sementara *RapidMiner* hanya mencapai *precision* 17.70%, *recall* 50%, dan *accuracy* 39.04%-44.57%. Namun, *recall* ulasan netral 0% akibat ketidakseimbangan dataset yang mengindikasikan potensi *overfitting*. Analisis menunjukkan ulasan positif menyoroti kualitas produk, sedangkan ulasan negatif terkait produk rusak dan ukuran tidak sesuai. Hasil ini memberi wawasan bagi penjual dalam meningkatkan produk dan layanan.

**Keywords:** *Natural Language Processing (NLP), Sentiment Analysis, Lexicon-Based, Multinomial Naïve Bayes, RapidMiner*

## 1. Pendahuluan

Marketplace adalah evolusi dari *e-commerce* yang berfungsi sebagai platform transaksi online antara penjual dan pembeli. Bukalapak, salah satu marketplace terbesar di Indonesia, memiliki 168,2 juta pengunjung sepanjang 2023 [1]. Platform ini banyak dimanfaatkan oleh produsen, UMKM, dan distributor untuk menjangkau calon pembeli. Salah satu kategori produk yang populer di Bukalapak adalah kemeja pria, yang kualitasnya dinilai berdasarkan foto, deskripsi, dan ulasan dari pembeli sebelumnya [2].

Dalam transaksi online, ulasan produk memegang peranan penting dalam membentuk reputasi suatu produk dan memengaruhi keputusan pembelian. Ulasan biasanya terdiri dari penilaian bintang dan komentar yang berisi apresiasi, kritik, serta saran dari pembeli. Produk dengan jumlah bintang tinggi umumnya lebih dipercaya dan menarik minat pembeli lain. Selain membantu calon pembeli dalam memilih produk, ulasan juga memberikan wawasan bagi penjual untuk meningkatkan kualitas produk dan layanan guna meningkatkan kepuasan pelanggan [3].

Meskipun jumlah bintang memberikan gambaran awal terhadap kualitas produk, komentar dalam ulasan sering kali mencakup lebih dari satu aspek penilaian, seperti harga, bahan, warna, bentuk, ukuran, kuantitas, pelayanan, dan kecepatan pengiriman [4]. Karena setiap ulasan hanya memiliki satu penilaian bintang, informasi penting dalam komentar tidak selalu dapat direpresentasikan secara akurat melalui jumlah bintang saja. Oleh karena itu, diperlukan analisis sentimen untuk memahami makna sebenarnya dari ulasan pembeli.

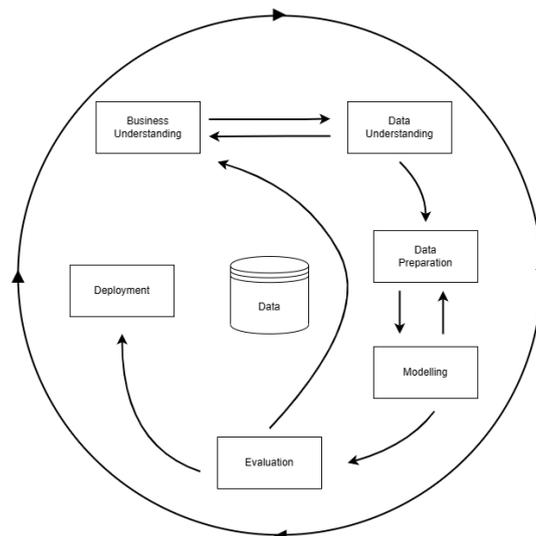
Analisis manual terhadap ribuan ulasan di marketplace tidak efisien dan rentan terhadap subjektivitas. Oleh sebab itu, penggunaan *Natural Language Processing* (NLP) menjadi penting untuk meningkatkan pemahaman terhadap isi komentar. NLP memungkinkan sistem untuk menormalkan teks, mengenali pola bahasa, dan menginterpretasikan sentimen dengan lebih akurat. Tanpa NLP, pengolahan data berbasis *Machine Learning* sulit mencapai hasil optimal.

Penelitian ini bertujuan untuk menganalisis sentimen ulasan produk kemeja pria di Bukalapak menggunakan pendekatan Lexicon-Based Approach dengan Indonesia Sentiment Lexicon (InSet Lexicon) sebagai referensi kata opini dalam bahasa Indonesia. Sentimen dalam ulasan diklasifikasikan ke dalam tiga kategori: positif, negatif, dan netral menggunakan algoritma Multinomial Naïve Bayes. Evaluasi model dilakukan dengan Confusion Matrix, menggunakan metrik akurasi, recall, dan precision.

Selain itu, hasil penelitian ini dibandingkan dengan analisis menggunakan RapidMiner untuk mengukur efektivitas model yang dikembangkan. Penelitian ini juga bertujuan untuk mengidentifikasi aspek utama dalam ulasan positif dan negatif, sehingga penjual dapat memahami faktor yang paling berpengaruh terhadap kepuasan pelanggan dan melakukan perbaikan kualitas produk serta layanan secara lebih tepat.

## 2. Metode Penelitian

Metode penelitian yang digunakan penulis adalah metode CRISP-DM. Metode tersebut diilustrasikan pada Gambar 1.



Gambar 1. CRISP-DM

### 2.1 Business Understanding

Berdasarkan latar belakang yang telah dipaparkan, penelitian ini berfokus pada dua permasalahan utama:

1. Pentingnya Ulasan Produk – Ulasan produk memiliki peran krusial bagi produsen dan konsumen, karena menjadi sumber utama informasi mengenai kualitas produk dan tingkat kepuasan pelanggan. Ulasan yang positif atau negatif dapat memengaruhi keputusan pembelian serta membangun reputasi merek di pasar digital.
2. Tantangan dalam Pengolahan Data – Volume ulasan yang besar dan beragam di platform e-commerce, seperti Bukalapak, menyulitkan analisis manual yang membutuhkan waktu lama dan kurang efisien. Dengan meningkatnya jumlah ulasan dan variasi produk, diperlukan pendekatan yang lebih efektif untuk menganalisis sentimen secara sistematis dan akurat.

Penelitian ini bertujuan untuk mengembangkan metode analisis sentimen yang dapat mengidentifikasi opini pengguna secara otomatis. Dengan pendekatan ini, pelaku bisnis dapat mengevaluasi kepuasan pelanggan dan kualitas produk berdasarkan sentimen positif, netral, dan negatif. Hasil analisis diharapkan memberikan wawasan yang lebih mendalam bagi penjual dalam merespons umpan balik pelanggan dengan strategi perbaikan atau peningkatan produk yang lebih tepat. Pemahaman yang lebih baik terhadap opini pelanggan juga dapat meningkatkan kepercayaan konsumen terhadap produk yang ditawarkan di marketplace.

### 2.2 Data Understanding

Dalam penelitian ini, pemilihan produk dilakukan berdasarkan hasil pencarian di Bukalapak dengan kata kunci "kemeja pria". Produk yang dipilih merupakan produk terlaris dengan beberapa kriteria:

1. Melikiki rating minimal bintang 3 keatas
2. Dijual oleh pelapak yang telah diverifikasi sebagai super seller oleh Bukalapak
3. Dipilih berdasarkan jumlah penjualan terbanyak, tanpa mempertimbangkan asal toko atau wilayah.

Alasan pemilihan kata kunci "kemeja pria" adalah sebagai berikut:

1. Berdasarkan Google Trends (Januari 2024), "kemeja pria" memiliki jumlah pencarian tertinggi dibandingkan dengan jenis kemeja lainnya di Indonesia, dengan skala relatif 100 sebagai kueri dengan frekuensi penelusuran tertinggi Dijual oleh pelapak yang telah diverifikasi sebagai super seller oleh Bukalapak [5].
2. Tingkat penjualan tinggi di Bukalapak, dengan produk terlaris mencapai 26 ribu unit terjual [6].
3. Memiliki fitur yang kompleks, termasuk model, bahan, warna, ukuran, potongan, dan jahitan, yang dapat berpengaruh terhadap variasi sentimen ulasan pelanggan.

Setelah menentukan produk, data dikumpulkan menggunakan metode *web scraping* pada situs Bukalapak dengan tools cURL PHP. Proses scraping terdiri dari beberapa tahap utama:

1. Mengakses URL produk menggunakan cURL.
2. Mengatur User-Agent agar dapat memperoleh data dalam format plain text HTML.
3. Mengekstrak public token untuk autentikasi API Bukalapak.
4. Memeriksa validitas public token. Jika kadaluarsa, proses kembali ke langkah (2).
5. Mengakses API Bukalapak ([api.bukalapak.com/reviews](https://api.bukalapak.com/reviews)) dan mengekstrak data ulasan.
6. Menyimpan data hasil ekstraksi ke dalam database.

Data yang diperoleh berjumlah 1.455 ulasan, yang kemudian diproses melalui tahap *pre-processing* dan dibagi menjadi lima skenario pembagian data latih dan data uji, yaitu: 90:10, 80:20, 70:30, 60:40, dan 50:50. (perbandingan data latih dan data uji). Tujuan dari pembagian ini adalah untuk mengetahui rasio data yang menghasilkan *precision*, *recall*, dan *accuracy* terbaik dalam evaluasi model.

Setelah data dibagi, dilakukan tahap *labeling* data pada latih. Meskipun rating bintang dapat memberikan indikasi awal, ulasan dapat mencakup lebih dari satu aspek, sehingga penting untuk menganalisis isi komentar secara menyeluruh [7].

Dalam penelitian ini, proses labeling dibantu secara otomatis menggunakan Indonesia Sentiment Lexicon (InSet Lexicon), yang diperoleh dari repositori GitHub [8]. InSet Lexicon terdiri dari:

- 3.609 kata positif, dengan bobot sentimen (+1 hingga +5).
- 6.609 kata negatif, dengan bobot sentimen (-1 hingga -5).

Polaritas suatu ulasan dihitung dengan menjumlahkan bobot dari setiap kata dalam ulasan yang tercantum di dalam *InSet Lexicon* (Indonesia *Sentiment Lexicon*). Prosedurnya adalah sebagai berikut [9].

- Hitung total skor positif dan negatif berdasarkan bobot kata yang terdapat dalam teks ulasan.
- Jumlahkan seluruh bobot untuk mendapatkan nilai polaritas akhir.

Rumus :

$$\text{Polaritas} = \sum_{i=1}^n \text{bobot}(kata_i) \quad (1)$$

Dimana :

- $kata_i$  = kata ke- $i$  dalam ulasan
- $\text{bobot}(kata_i)$  = nilai sentimen dari  $kata_i$  berdasarkan InSet Lexicon

- Interpretasi hasil:
  - Jika polaritas > 0, ulasan dikategorikan sebagai positif.
  - Jika polaritas < 0, ulasan dikategorikan sebagai negatif.
  - Jika polaritas = 0, ulasan dikategorikan sebagai netral.

### 2.3 Data Preparation

Data latih dan data uji diolah dalam proses *pre-processing* menggunakan pendekatan *Natural Language Processing* (NLP) untuk memperbaiki bahasa dalam ulasan. Tahapan *pre-processing* ini

mencakup pembersihan data (*cleaning*), *case folding*, normalisasi, *Tokenizing*, *stopword removal*, dan *stemming*. Proses ini membantu sistem memahami bahasa tidak baku dan singkatan yang sering muncul dalam ulasan pengguna.

1. *Cleaning* : menghapus karakter yang tidak relevan atau tidak memberikan kontribusi signifikan terhadap analisis sentimen. Beberapa karakter tersebut termasuk angka, tanda baca, dan karakter khusus yang tidak memiliki nilai informatif dalam pemahaman sentimen ulasan pengguna *E-Commerce*.
2. *Case Folding* : mengubah semua huruf kapital menjadi huruf kecil. Hal ini dilakukan karena tidak semua data teks konsisten dalam penggunaan huruf kapital.
3. Normalisasi : memperbaiki kata-kata dalam ulasan agar menghasilkan kalimat yang baik dan benar sesuai dengan kaidah tata bahasa Indonesia.
4. *Tokenizing* : memecah kalimat menjadi unit-unit kata.
5. *Stopword Removal* : *Stop Word* merupakan kelompok kata yang tidak memiliki hubungan signifikan dengan inti pembicaraan atau subyek yang sedang dibahas. Kata-kata ini umumnya terdiri dari kata imbuhan, kata sambung, dan lain sebagainya
6. *Stemming* : mengubah kata-kata dalam dokumen teks menjadi kata dasar atau akar kata.

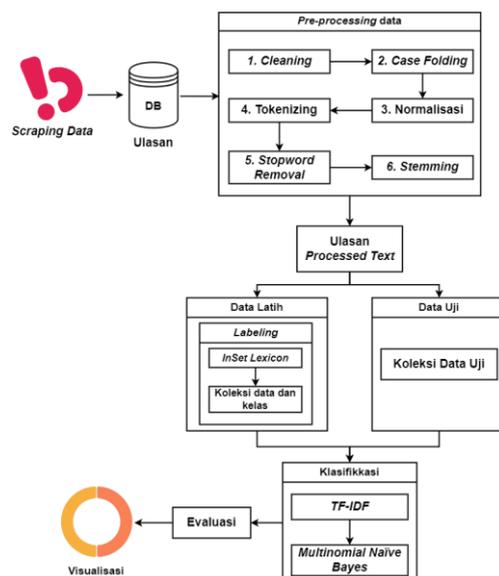
Tabel berikut menunjukkan tahapan *pre-processing* yang dilakukan dalam penelitian, beserta jumlah kata yang terpengaruh dan contoh transformasi dari teks mentah menjadi bentuk siap analisis.

**Tabel 1.** *Pre-processing*

Tahapan <i>Pre-Processing</i>	Jumlah Kata Terpengaruh	Contoh Sebelum	Contoh Sesudah
<i>Cleaning</i>	532 kata	"mantap!!!", "cepat...", "bagus#1", "mantul~"	"mantap", "cepat", "bagus", "mantul"
<i>Case Folding</i>	1.455 ulasan	"Bagus", "CEPAT", "MAHAL"	"bagus", "cepat", "mahal"
Normalisasi	217 kata	"bgs", "manteb", "brng", "gk", "trima kasih"	"bagus", "mantap", "barang", "tidak", "terima kasih"
<i>Tokenizing</i>	1.455 kalimat	"pengiriman cepat dan rapi"	["pengiriman", "cepat", "dan", "rapi"]
<i>Stopword Removal</i>	863 kata	"dan", "itu", "ke", "untuk"	(dihapus dari teks)
<i>Stemming</i>	645 kata	"membeli", "terjual", "pengiriman", "pelayanan", "dibayar"	"beli", "jual", "kirim", "layan", "bayar"

## 2.4 Modeling

Dalam penelitian ini model yang dibuat adalah untuk mengevaluasi kualitas produk dan kepuasan pelanggan dengan menggunakan pendekatan analisis sentimen. Penerapan model analisis sentimen pada penelitian ini dapat dilihat pada Gambar 2.



**Gambar 2.** Model Analisis Sentimen

Pada Gambar 2 menunjukkan model analisis sentimen yang diterapkan dalam penelitian ini. Proses dimulai dengan scraping data dari *e-commerce* Bukalapak menggunakan API untuk mengumpulkan ulasan produk. Data yang dikumpulkan (nama pemberi ulasan, title, dan konten) disimpan dalam database. Data mentah ini kemudian melalui tahap *pre-processing*, yang mencakup *cleaning*, *case folding*, normalisasi, *tokenizing*, *stopword removal*, dan *stemming*. Setelah *pre-processing*, data dibagi menjadi 5 skenario percobaan (perbandingan data latih dan uji: 90:10, 80:20, 70:30, 60:40, dan 50:50) untuk mengukur precision, recall, dan accuracy. *Labeling* otomatis dilakukan menggunakan *InSet Lexicon* untuk menentukan sentimen ulasan (positif, netral, atau negatif). klasifikasi dilakukan dengan menggunakan TF-IDF (Term Frequency - Inverse Document Frequency) untuk menghitung pentingnya kata dalam dokumen. Dengan rumus sebagai berikut[10].

TF-IDF sendiri pada dasarnya terbagi menjadi 2 istilah *Term Frequency* dan *Inverse Document Frequency*.

*Term Frequency* (TF) menghitung jumlah kemunculan suatu kata dalam dokumen.

$$tf_{i,j} = \frac{n_{i,j}}{\sum_k n_{k,j}} \quad (2)$$

- $n_{i,j}$  adalah angka kejadian pada kata di dalam sebuah file
- $\sum_k n_{k,j}$  adalah jumlah dari kejadian pada semua kata yang ada di dalam sebuah file

Sedangkan *Inverse Document Frequency* (IDF) digunakan untuk menghitung tingkat kepentingan suatu istilah. Contoh nya adalah “yang”, “di”, “ke” atau kata-kata yang digunakan sebagai preposisi dan konjungsi. Istilah-istilah tersebut merupakan istilah yang sering muncul namun tidak memiliki arti yang begitu penting.

$$idf_i = \log \frac{|D|}{|\{j: t_i \in d_j\}|} \quad (3)$$

- $|D|$  adalah jumlah total file di dalam korpus
- $|\{j: t_i \in d_j\}|$  adalah jumlah dokumen yang mengandung huruf  $t_i$ .

Apabila kata tersebut tidak ada di dalam korpus maka akan menyebabkan pembagi nya menjadi nol. Oleh sebab itu maka digunakanlah rumus seperti berikut  $1 + |\{j: t_i \in d_j\}|$ . Maka dari kedua rumus diatas dapat kita pahami, semakin sering atau semakin tinggi kemunculan suatu kata dalam sebuah dokumen maka akan memberikan frekuensi yang tinggi terhadap kata tersebut dan semakin jarang kata tersebut atau sedikit tingkat kemunculan kata tersebut maka akan menghasilkan nilai yang tinggi atau dalam arti lain kata tersebut memiliki makna yang penting dalam suatu dokumen.

$$W_{i,j} = t f_{i,j} x \left( \frac{D}{D f_j} \right) \quad (4)$$

Selain itu, dilakukan klasifikasi menggunakan *Multinomial Naïve Bayes*. Dengan rumus sebagai berikut [11].

$$P(w_i, c_j) = \frac{\text{count}(w_i, c_j) + 1}{\sum \text{count}(w_i, c_j) + |V|} \quad (5)$$

Pada perhitungan likelihood, digunakan pendekatan *Laplace smoothing* untuk menghindari probabilitas 0. Penjelasan dari variabel dalam rumus tersebut adalah sebagai berikut:

- $P(w_i, c_j)$  adalah probabilitas kemunculan kata  $w_i$  pada kelas  $c_j$ .
- $\text{count}(w_i, c_j)$  adalah jumlah kemunculan kata  $w_i$  pada kelas  $c_j$ .
- $|V|$  adalah jumlah kata unik pada seluruh kelas.

Hasil klasifikasi dievaluasi menggunakan *Confusion Matrix Multi-Class* dan dibandingkan dengan platform RapidMiner. Sistem yang dikembangkan juga menyajikan visualisasi berupa word cloud dan bar chart untuk menggambarkan distribusi sentimen serta kinerja model.

### 2.5 Evaluation

Pada tahap evaluasi, model klasifikasi *Multinomial Naïve Bayes* digunakan untuk menganalisis sentimen ulasan produk. Pengujian dilakukan dengan membandingkan prediksi model terhadap data uji dan data latih yang dilabel otomatis menggunakan kamus InSet untuk mengategorikan sentimen menjadi positif, negatif, dan netral. Data latih dan data uji dibagi dalam lima skenario (90:10, 80:20, 70:30, 60:40, dan 50:50) untuk menentukan rasio terbaik berdasarkan *precision*, *recall*, dan *accuracy*. Evaluasi kinerja model menggunakan Confusion Matrix, yang merupakan tabel 3x3 untuk tiga kelas sentimen. Confusion Matrix ini mencakup *True Positive* (TP), *True Negative* (TN), *True Neutral* (TNt), *False Positive* (FP), *False Negative* (FN), dan *False Neutral* (FNt), membantu memvisualisasikan akurasi prediksi model terhadap hasil sebenarnya [12].

**Tabel 2. Confusion Matrix Multi-Class**

		Predicted Class		
		Positive	Negative	Neutral
True Class	Positive	True Positive (TP)	False Negative (FN)	False Neutral (FNt)
	Negative	False Positive (FP)	True Negative (TN)	False Neutral (FNt)
	Neutral	False Positive (FP)	False Negative (FN)	True Neutral (TNt)

Untuk mengetahui rata-rata *precision* dan *recall* dalam penelitian ini, juga digunakan *Weighted Average*. *Weighted Average* menghitung setiap Matrik dengan mempertimbangkan jumlah sampel yang ada di setiap kelas. Hal ini membantu memberikan gambaran yang lebih akurat tentang kinerja model terhadap masing-masing kelas, memperhitungkan perbedaan jumlah sampel yang direpresentasikan oleh setiap kelas dalam evaluasi [13].

$$\text{Weighted Average Precision} = \sum_i \left( \frac{\text{Total Kelas } i}{\text{Total Keseluruhan}} \times \text{Precision Kelas } i \right) \quad (6)$$

$$\text{Weighted Average Recall} = \sum_i \left( \frac{\text{Total Kelas } i}{\text{Total Keseluruhan}} \times \text{Recall Kelas } i \right) \quad (7)$$

## 2.6 Deployment

Tahapan perencanaan penyebaran sistem ini didasarkan pada rancangan yang telah disiapkan sebelumnya. Hal ini bertujuan untuk membantu visualisasi hasil penelitian. Sistem ini dikembangkan sebagai web-site menggunakan bahasa pemrograman PHP dengan *framework* Laravel. Antarmuka pengguna dibangun dengan menggunakan HTML, CSS, dan JavaScript. Mulai dari tahap pengambilan data dari Bukalapak menggunakan cURL PHP hingga tahap visualisasi pada sistem, semua proses dilakukan dengan menggunakan bahasa pemrograman PHP dan MySQL sebagai basis data yang dikelola.

## 3. Hasil dan Pembahasan

Pengujian dilakukan menggunakan metode *Confusion Matrix Multi-Class* karena terdapat lebih dari dua kelas, yaitu kelas positif, negatif, dan netral. Pengujian ini menggunakan sampel data yang diambil dari dataset ulasan produk murah - kemeja batik termurah - kemeja pria lengan pendek termurah dari toko BATIK MURAH JAYA, sebanyak 1.455 data komentar ulasan. Data ini dibagi menjadi lima skenario percobaan dengan perbandingan data latih dan data uji sebagai berikut: 90:10, 80:20, 70:30, 60:40, dan 50:50.

### 3.1. Hasil Pengujian Data Uji

Kinerja model dievaluasi dengan menghitung *confusion matrix* untuk setiap skenario pembagian data. Setiap *matrix* menyediakan informasi yang diperlukan untuk menghitung precision, recall, dan accuracy, yang kemudian dibandingkan antara skenario yang berbeda untuk menentukan pembagian data yang optimal.

**Tabel 2.** Pembagian Data Uji dan Data Latih

Proposi	Sistem		RapidMiner	
	Data Latih	Data Uji	Data Latih	Data Uji
90:10	1310	145	1309	146
80:20	1164	291	1164	291
70:30	1019	436	1018	437
60:40	873	582	873	582
50:50	728	727	728	727

**Tabel 3.** Hasil Evaluasi Sistem yang Dikembangkan

proposisi	Precision			Recall			accuracy
	Positif	Negatif	Netral	Positif	Negatif	Netral	
90:10	93%	75%	100%	97%	72%	0%	91%
80:20	94%	77%	100%	97%	74%	0%	92%
70:30	94%	75%	100%	97%	72%	0%	92%
60:40	94%	69%	100%	97%	59%	0%	92%
50:50	94%	92%	100%	99%	70%	0%	94%

Sistem yang dikembangkan menunjukkan performa tinggi dalam mendeteksi ulasan positif, dengan precision antara 93%-94% dan recall hingga 99% pada skenario 50:50. Precision untuk kelas negatif bervariasi antara 69%-92%, menunjukkan bahwa model cukup akurat dalam mengidentifikasi ulasan positif dan negatif. Namun, model kesulitan mendeteksi ulasan netral, dengan recall untuk kelas netral selalu 0% akibat ketidakseimbangan data, di mana ulasan positif jauh lebih banyak daripada netral. Hal ini menyebabkan model cenderung mengklasifikasikan ulasan sebagai positif atau negatif, mengabaikan ulasan netral. Kesulitan dalam mendeteksi ulasan netral mungkin disebabkan oleh fitur yang kurang menonjol atau mirip dengan ulasan positif/negatif, sehingga sulit dibedakan oleh model. Meskipun *accuracy* keseluruhan tetap tinggi (91%-94%), rendahnya recall pada kelas netral menunjukkan indikasi *overfitting*, di mana model kurang mampu menangani kelas dengan representasi rendah.

**Tabel 4.** Hasil Evaluasi RapidMiner

proposisi	Precision			Recall			accuracy
	Positif	Negatif	Netral	Positif	Negatif	Netral	
90:10	13.64%	78.57%	39.47%	34.62%	32.67%	78.95%	39.04%
80:20	17.70%	82.93%	33.33%	39.22%	33.66%	84.21%	41.24%
70:30	16.00%	82.58%	47.50%	46.75%	35.97%	66.67%	41.88%
60:40	17.17%	83.06%	55.88%	50.00%	37.62%	75.00%	44.57%
50:50	14.24%	85.56%	38.78%	38.28%	31.68%	80.85%	39.20%

Performa RapidMiner lebih rendah dibandingkan sistem yang dikembangkan. Precision untuk kelas positif sangat rendah, hanya 13,64% pada skenario 90:10, meskipun ada perbaikan di skenario lain. Recall untuk kelas positif juga kurang memuaskan, dengan nilai tertinggi hanya 50% pada skenario 60:40. Meskipun recall untuk kelas netral lebih tinggi, accuracy keseluruhan RapidMiner tetap rendah, hanya 44,67%.

Penelitian ini menerapkan *Weighted Average* dalam perhitungan rata-rata hasil *precision* dan *recall* pada hasil sistem yang dikembangkan dan hasil platform RapidMiner pada masing- masing skenario pembagian data.

**Tabel 5.** Hasil Weighted Avg Sistem

Proporsi	Precision	Recall	Accuracy
90:10	83%	86%	91%
80:20	86%	88%	92%
70:30	87%	90%	92%
60:40	85%	87%	92%
50:50	88%	91%	94%

Hasil *weighted average* untuk sistem yang dikembangkan menunjukkan precision antara 83% hingga 88%, recall antara 86% hingga 91%, dan accuracy mencapai 94% pada proporsi 50:50. Sistem ini menunjukkan performa konsisten yang baik, terutama dalam mendeteksi ulasan positif. Model berfungsi optimal dengan proporsi data pelatihan lebih besar, memungkinkan klasifikasi yang lebih akurat.

**Tabel 6.** Hasil Weighted Avg RapidMiner

Proporsi	Precision	Recall	Accuracy
90:10	23%	18%	39%
80:20	24%	20%	41.27%
70:30	26%	19%	41.88%
60:40	29%	22%	44.67%
50:50	24%	18%	39.20%

Hasil *weighted average* untuk RapidMiner menunjukkan *precision* antara 23% hingga 29%, *recall* antara 18% hingga 22%, dan *accuracy* tertinggi 44.67% pada skenario 60:40. Ini menunjukkan keterbatasan RapidMiner dalam mendeteksi ulasan, terutama ulasan positif, dengan precision dan recall yang rendah di semua skenario.

### 3.2. Hasil Evaluasi

Grafik hasil prediksi data uji menunjukkan distribusi prediksi model klasifikasi sentimen ke dalam tiga kategori: positif, negatif, dan netral. Grafik ini didasarkan pada skenario 50:50 yang menunjukkan performa terbaik dari sistem. Dari hasil yang ditampilkan, prediksi sentimen positif mendominasi dengan jumlah 627, yang digambarkan oleh batang biru yang tinggi di grafik.



Gambar 3. Word Cloud Positif

### 3.3. Hasil Visualisasi Wordcloud

Secara keseluruhan, *wordcloud* positif menunjukkan perasaan-perasaan positif yang dapat disebabkan oleh berbagai faktor, seperti kebahagiaan atas suatu pencapaian, rasa puas dengan suatu situasi, atau sikap positif dalam menghadapi hidup. Pada ulasan dengan label positif, beberapa kata muncul dengan frekuensi tinggi, menunjukkan aspek-aspek yang dihargai oleh konsumen. Kata "bagus" muncul sebanyak 1100 kali, "cepat" muncul 172 kali, "puas" muncul 201 kali, dan "rapi" muncul 54 kali.



Gambar 4. Wordcloud Positif

*Wordcloud* negatif menunjukkan kumpulan kata-kata yang memiliki makna negatif. Kata-kata ini dapat digunakan untuk mengungkapkan perasaan atau pikiran yang tidak menyenangkan, seperti kekecewaan, kemarahan, dan ketakutan. Dalam ulasan dengan label negatif, beberapa kata muncul dengan frekuensi tinggi, menunjukkan aspek-aspek yang menjadi keluhan konsumen. Kata "rusak" muncul sebanyak 121 kali, "kecil" muncul 60 kali, "ragu" muncul 51 kali, dan "lama" muncul 57 kali. Penggunaan kata "lama" dan "ragu" menunjukkan bahwa pembeli menerima produk yang tidak sesuai dengan harapan mereka.



Gambar 5. Wordcloud Negatif

*Wordcloud* netral menunjukkan kata-kata yang umum digunakan dalam komunikasi sehari-hari, transaksi, atau penyampaian informasi objektif tanpa sentimen emosional yang kuat. Kata-kata seperti "tidak," "biasa," "saya," dan "puas" sering muncul, mengindikasikan ulasan netral yang biasanya berupa pertanyaan atau pernyataan tanpa opini negatif maupun positif.



Gambar 6. Wordcloud Netral

#### 4. Conclusion

Berdasarkan hasil penelitian yang dilakukan dan pengujian tingkat akurasi yang didapatkan dapat ditarik kesimpulan sebagai berikut :

1. Penelitian ini menegaskan pentingnya *pre-processing* data dalam analisis sentimen untuk memastikan hasil yang konsisten dan akurat. Tahapan seperti *cleaning*, normalisasi, dan *stopword removal* membantu menyelaraskan variasi penulisan dalam ulasan produk, termasuk singkatan seperti "bgs" menjadi "bagus" dan penulisan huruf ganda seperti "ceepaat" menjadi "cepat." Beberapa kata penting seperti "wah," "tidak," dan "oke" dipertahankan karena relevansinya dalam Indonesia Sentiment Lexicon, yang berperan penting dalam mendukung analisis lebih mendalam terhadap ulasan.
2. Sistem yang dikembangkan menunjukkan performa unggul dibandingkan RapidMiner dalam mengklasifikasikan sentimen. Pada pembagian data 70:30, 60:40, dan 50:50, sistem mencapai precision hingga 94%, recall hingga 99%, dan accuracy maksimal 94%. Meski begitu, tantangan muncul pada ulasan netral, di mana recall bernilai 0% akibat *imbalanced dataset*, menunjukkan adanya potensi *overfitting*. Sebaliknya, RapidMiner hanya menghasilkan accuracy 39.04%-44.57% dengan precision tertinggi 17.70%, menunjukkan sistem yang dikembangkan lebih andal.
3. Analisis barchart dan wordcloud mendukung temuan ini, di mana ulasan positif mendominasi (627 ulasan), mencerminkan kepuasan pada kualitas produk, pengiriman, dan pengemasan. Ulasan negatif (89 ulasan) mengungkapkan masalah terkait produk rusak dan pengiriman lambat, sedangkan ulasan netral (14 ulasan) menggunakan kata-kata tanpa sentimen kuat. Keseluruhan temuan ini memberikan wawasan berharga bagi penjual untuk memperbaiki kualitas produk, layanan, dan pengiriman, sehingga dapat mempertahankan ulasan positif dan mengurangi keluhan pelanggan. Integrasi pre-processing data yang baik, kinerja model yang unggul, dan analisis hasil ulasan secara visual memberikan panduan strategis untuk meningkatkan pengalaman pelanggan.

#### References

- [1] A. Ahdiat, "5 E-Commerce dengan Pengunjung Terbanyak Sepanjang 2023," databoks. Accessed: Feb. 02, 2024. [Online]. Available: <https://databoks.katadata.co.id/datapublish/2024/01/10/5-e-commerce-dengan-pengunjung-terbanyak-sepanjang-2023>
- [2] H. Wang and Y. Wang, "A Review of Online Product Reviews," *Journal of Service Science and Management*, vol. 13, no. 01, pp. 88–96, 2020, doi: 10.4236/jssm.2020.131006.
- [3] I. Servanda Ratu, P. Sari Kemala Reno, and N. Ananda Adhitya, "Peran Ulasan Produk Dan Fot Produk Yang Ditampilkan Penjual Pada Marketplace Shopee Terhadap Minat Beli Pria Dan Wanita," *JURNAL MANAJEMEN DAN BISNIS*, vol. 2, no. 2, 2019, [Online]. Available: <http://jurnal.uts.ac.id>
- [4] R. Maulana Arrasyid, D. Enggar Putera, and A. Yunizar Pratama Yusuf, "Analisis Sentimen Review Pembelian Produk di Marketplace Shopee Menggunakan Pendekatan Natural Language Processing," *Jurnal TEKNO KOMPAK*, vol. 18, no. 2, pp. 319–330, 2024.
- [5] Google, "Google Trends 'kemeja,'" Google. Accessed: May 06, 2024. [Online]. Available: <https://trends.google.co.id/trends/explore?geo=ID&q=kemeja>
- [6] Bukalapak, "Produk Kemeja Pria diBukalapak," Bukalapak. Accessed: Feb. 01, 2024. [Online]. Available:

- [https://www.bukalapak.com/products?search%5Bkeywords%5D=kemeja%20pria&search%5Brating\\_gte%5D=3&search%5Bsort\\_by%5D=weekly\\_sales\\_ratio%3Adesc&search%5Btop\\_seller%5D=1](https://www.bukalapak.com/products?search%5Bkeywords%5D=kemeja%20pria&search%5Brating_gte%5D=3&search%5Bsort_by%5D=weekly_sales_ratio%3Adesc&search%5Btop_seller%5D=1)
- [7] E. Muktafin Hari, K. Kusri, and E. Luthfi Taufiq, "Analisis Sentimen pada Ulasan Pembelian Produk di Marketplace Shopee Menggunakan Pendekatan Natural Language Processing," *Jurnal Eksplora Informatika*, vol. 10, no. 1, pp. 32–42, Sep. 2020, doi: 10.30864/eksplora.v10i1.390.
- [8] A. Hendra, "Kamus InSet (Indonesia Sentiment Lexicon)," GitHub. Accessed: Dec. 13, 2023. [Online]. Available: <https://github.com/agushendra7/twitter-sentiment-analysis-using-inset-and-random-forest>
- [9] D. Musfiroh, U. Khaira, P. E. P. Utomo, and T. Suratno, "Analisis Sentimen terhadap Perkuliahan Daring di Indonesia dari Twitter Dataset Menggunakan InSet Lexicon," *MALCOM*, vol. 1, no. 1, pp. 24–33, 2021.
- [10] L. Siniwi Maharani, A. Prahutama, and A. Hakim Rachman, "Query Expansion Ranking pada Analisis Sentimen Menggunakan Klasifikasi Multinomial Naïve Bayes," *JURNAL GAUSSIAN*, vol. 10, no. 3, pp. 377–387, 2021.
- [11] N. Hidayah and Dodiman, "Implementasi Algoritma Multinomial Naïve Bayes, TF-IDF dan Confusion Matrix dalam Pengklasifikasian Saran Monitoring dan Evaluasi Mahasiswa Terhadap Dosen Teknik Informatika Universitas Dayanu Ikhsanuddin," *Jurnal Akademik Pendidikan Matematika*, vol. 10, no. 1, pp. 8–15, 2024, doi: 10.55340/japm.v10i1.1491.
- [12] S. Anggina, N. Y. Setiawan, and F. A. Bachtiar, "Analisis Ulasan Pelanggan Menggunakan Multinomial Naïve Bayes Classifier dengan Lexicon-Based dan TF-IDF Pada Formaggio Coffee and Resto," *is The Best Accounting Information Systems and Information Technology Business Enterprise this is link for OJS us*, vol. 7, no. 1, pp. 76–90, Sep. 2022, doi: 10.34010/aisthebest.v7i1.7072.
- [13] Yuyun, N. Hidayah, and S. Sahibu, "Algoritma Multinomial Naïve Bayes Untuk Klasifikasi Sentimen Pemerintah Terhadap Penanganan Covid-19 Menggunakan Data Twitter," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 5, no. 4, pp. 820–826, Aug. 2021, doi: 10.29207/resti.v5i4.3146.

*This page is intentionally left blank.*