

# Analisis Sentimen *Review* Pembelian Produk di *Marketplace* Shopee Menggunakan Pendekatan *Natural Language Processing*

Rizky Maulana Arrasyid<sup>1</sup>, Diaz Enggar Putera<sup>2</sup>, Ajif Yunizar Pratama Yusuf<sup>3,\*</sup>

<sup>1,2,3</sup> Ilmu Komputer, Universitas Bhayangkara Jakarta Raya  
Jl. Raya Perjuangan No.81, RT.003/RW.002, Marga Mulya, Kec. Bekasi Utara, Kota Bekasi, Jawa Barat

Email: <sup>1</sup>202010225308@mhs.ubharajaya.ac.id , <sup>2</sup>202010225016@mhs.ubharajaya.ac.id

<sup>3,\*</sup>ajif.yunizar@dsn.ubharajaya.ac.id

<sup>\*)</sup> Email Penulis Utama

**Abstrak**—Penelitian ini menganalisis ulasan produk yang ada di pasaran, dengan fokus pada produk "Kaos Oversize", untuk mengklasifikasikannya ke dalam ulasan positif dan negatif. Penelitian ini bertujuan untuk menunjukkan keefektifan penggunaan algoritma *K-Nearest Neighbors* (KNN) dan *Term Frequency-Inverse Document Frequency* (TF-IDF) dengan pendekatan *Natural Language Processing* (NLP) dalam mengklasifikasikan ulasan produk. Penelitian ini menemukan bahwa metode NLP mencapai tingkat akurasi, presisi, dan recall yang lebih tinggi dibandingkan dengan tidak menggunakan NLP. Hasil penelitian menunjukkan bahwa menganalisis kata kunci dalam ulasan dapat mewakili opini keseluruhan pembeli terhadap produk, yang dapat menjadi informasi yang berguna bagi pengecer untuk mengevaluasi produk dan layanan mereka.

**Kata Kunci:** Analisis sentimen, Ulasan produk, *Marketplace*, KNN, NLP

**Abstract**— Abstract This study analyzes product reviews in the market, focusing on the "Kaos Oversize" product, to classify them into positive and negative reviews. The research aims to demonstrate the effectiveness of using *K-Nearest Neighbors* (KNN) and *Term Frequency-Inverse Document Frequency* (TF-IDF) algorithms with *Natural Language Processing* (NLP) approach in classifying product reviews. The study found that the NLP method achieved a higher accuracy, precision, and recall rate compared to not using NLP. The results suggest that analyzing keywords in reviews can represent the overall opinion of the buyers towards the product, which can be useful information for retailers to evaluate their products and services.

**Keywords:** Sentiment analysis, Product reviews, *Marketplace*, KNN, NLP.

## 1. PENDAHULUAN

*Marketplace* adalah sebuah wadah pemasaran produk secara elektronik yang mempertemukan banyak penjual dan pembeli untuk saling bertransaksi[1]. Hasil pendataan survei eCommerce sampai dengan 15 September 2022 menunjukkan hanya 34,10 persen usaha yang melakukan kegiatan eCommerce. Hal ini menunjukkan bahwa usaha yang menerima pesanan atau melakukan penjualan barang/jasa melalui internet di Indonesia masih tergolong rendah dan masih didominasi dengan jenis usaha konvensional. Informasi lain yang diperoleh dari Survei eCommerce 2022 adalah sebaran usaha. Sama dengan hasil survei tahun lalu, usaha eCommerce masih terpusat di Pulau Jawa. Pada tahun 2021, dari 2.868.178 usaha eCommerce, sebanyak 1.497.655 usaha (52,22 persen) berlokasi di pulau terpadat di Indonesia. Hal ini dapat dimaklumi mengingat pangsa pasar terbesar, sentra produksi dan infrastruktur pendukung koneksi internet yang memadai berlokasi di pulau Jawa[2].

*Marketplace* sebagai salah satu instrument untuk menaikkan daya beli masyarakat, pada akhirnya akan menjadi katup pengaman dari situasi krisis moneter. Pengembangan *marketplace* menjadi sangat strategis dalam menggerakkan perekonomian nasional, mengingat kegiatan usahanya mencakup hampir semua lapangan usaha sehingga kontribusi *marketplace* menjadi sangat besar bagi peningkatan pendapatan[3]. Untuk memperluas pasar dapat dilakukan melalui keterlibatan para pelaku usaha dalam e-commerce. Saat ini pertumbuhan e-commerce cukup signifikan karena gaya masyarakat yang mulai bergeser dari belanja *offline* menjadi cenderung berbelanja *online*. Pergeseran gaya belanja masyarakat, ditandai juga dengan pertumbuhan *marketplace* e-commerce yang tersedia[4]. Hal ini menjadi peluang bagi para pelaku usaha kecil untuk dapat terlibat dalam maraknya pilihan *marketplace* e-commerce yang dapat diakses secara gratis. Melansir situs goukm.id, keterlibatan pelaku usaha kecil dalam pasar online dapat meningkatkan pertumbuhan ekonomi di Indonesia. Selain itu Sentimen dari komentar yang diberikan pelanggan terhadap suatu produk dapat mencerminkan seberapa puasnya pelanggan terhadap produk tersebut. Sehingga dapat diketahui seberapa besar kepuasan pelanggan setelah atau selama menggunakan produk tersebut[5].

Dalam konteks ini, analisis sentimen dari ulasan pembelian produk menjadi penting untuk dipelajari. Dengan menggunakan teknik *Natural Language Processing* (NLP), kita dapat menggali wawasan yang berharga dari sekumpulan ulasan, memungkinkan pengecer dan penjual untuk memahami dan merespons dengan lebih baik terhadap kebutuhan dan preferensi pelanggan. Pada penelitian ini peneliti melakukan analisis sentimen pada ulasan produk di *marketplace* Shopee. Bagian ulasan produk terdiri dari isi komentar dengan format teks bebas dan peringkat jumlah bintang dari 1 sampai 5. Isi komentar ulasan digunakan untuk mengetahui informasi yang menjadi fokus pembeli dalam memberikan ulasan. Isi komentar ulasan dapat berisi lebih dari satu penilaian atribut produk, sedangkan setiap ulasan hanya memiliki satu penilaian jumlah bintang, sehingga jumlah bintang tidak dapat mewakili setiap fitur produk yang dinilai oleh pembeli. Informasi yang disampaikan pembeli dapat merujuk pada fitur produk seperti harga, kualitas, bahan, warna, bentuk, ukuran, rasa, jumlah, maupun pada pelayanan yang diberikan seperti pengemasan, lama pengiriman, dan tanggapan penjual[6]. Dan dalam penelitian ini, kami bertujuan untuk melakukan analisis sentimen terhadap ulasan pembelian produk di *marketplace* Shopee. Kami akan menerapkan pendekatan NLP untuk mengklasifikasikan ulasan menjadi kategori sentimen yang sesuai, seperti sentiment positif, dan negatif. Dengan demikian, kami berharap dapat memberikan pemahaman yang lebih dalam persepsi pengguna terhadap produk di Shopee dan kontribusi terhadap pengembangan strategi pemasaran dan penjualan yang lebih efektif dalam konteks e-commerce[7].

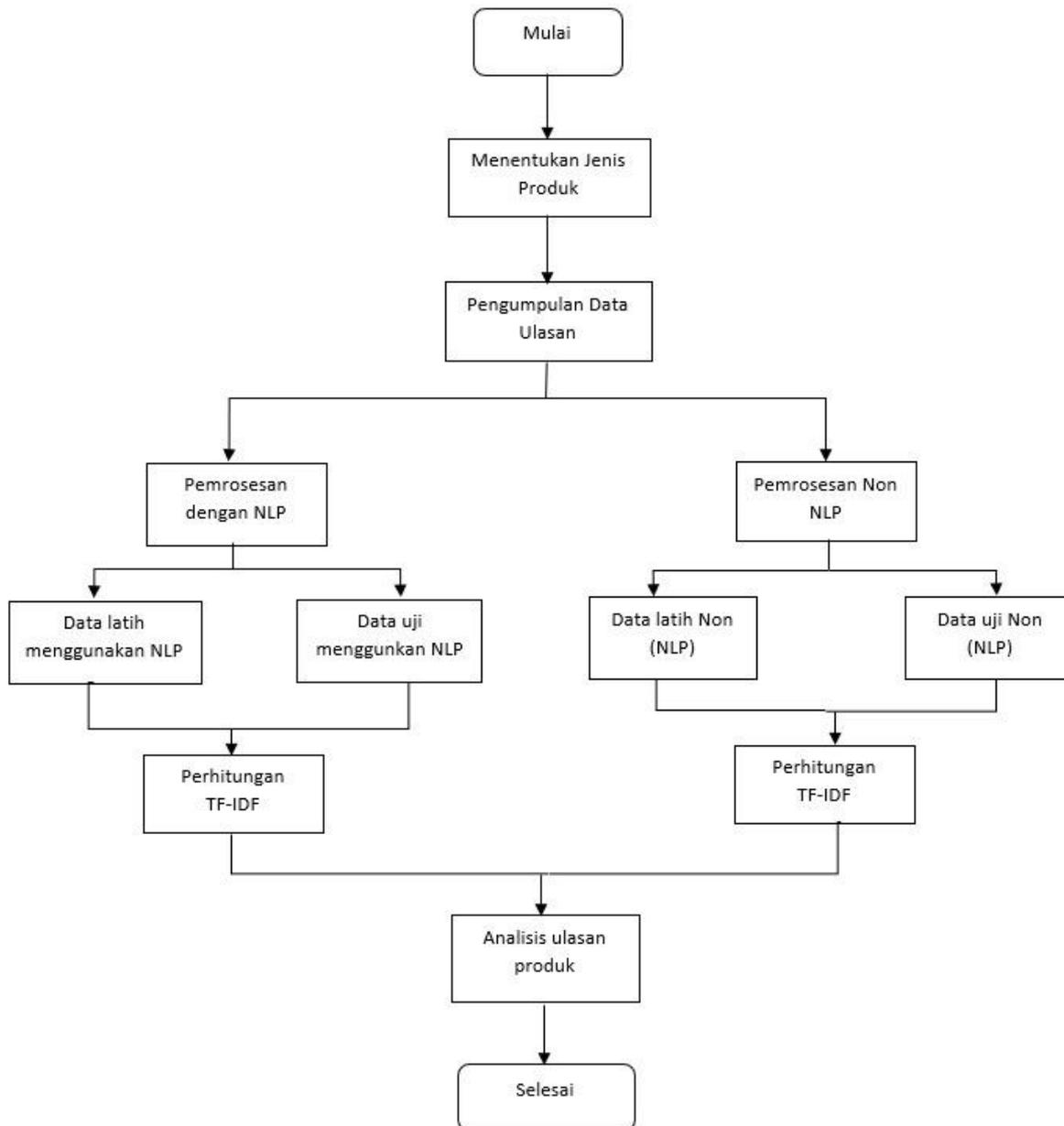
## 2. METODE PENELITIAN

### 2.1 Pengumpulan Data

Jenis survei ini memakai teknik kualitatif, & pengumpulan data buat survei ini memakai observasi & studi pustaka. Metode analisis yg dipakai pada penelitian ini merupakan Menjelaskan Aliran Pencarian dalam Analisis Sentimen Ulasan Produk Shopee menggunakan *Natural Language Processing* (NLP)[8].

### 2.2 *Natural Language Processing*

*Natural Language Processing* (NLP) adalah cabang dari ilmu komputer dan kecerdasan buatan yang berfokus pada interaksi antara komputer dan bahasa manusia alami. Tujuan utama NLP adalah untuk mengizinkan komputer untuk memahami, menganalisis, dan menghasilkan bahasa manusia secara efektif. Hal ini melibatkan berbagai teknik dan algoritma untuk memproses, memahami, dan menghasilkan teks dan ucapan manusia.[9]. Penelitian yang digambarkan dalam Gambar 1 dimulai dengan tahap identifikasi produk yang akan dianalisis. Langkah berikutnya melibatkan pengumpulan data ulasan produk terkait. Selanjutnya, dilakukan tahap pra-pemrosesan pada ulasan yang telah terkumpul dengan menggunakan dua pendekatan yang berbeda, yaitu dengan dan tanpa pendekatan NLP. Hasil dari proses ini menghasilkan dua dataset yang berbeda: satu dataset menggunakan pendekatan NLP, sementara dataset lainnya tidak menggunakan pendekatan NLP. Kemudian, setiap dataset dibagi menjadi dua bagian, yaitu data latih dan data uji. Selanjutnya, diterapkan perhitungan pembobotan *Term Frequency-Inverse Document Frequency* (TF-IDF) pada kedua dataset untuk kemudian dibandingkan hasilnya. Metode pembobotan TF-IDF digunakan untuk mengevaluasi pentingnya kata-kata dalam sebuah dokumen berdasarkan frekuensi kemunculan kata tersebut dalam dokumen tersebut dan frekuensi kemunculan kata tersebut dalam seluruh kumpulan dokumen.[10]. Hasil dari perbandingan antara kedua dataset yang telah melalui proses pembobotan TF-IDF kemudian dianalisis untuk mengevaluasi efektivitas pendekatan NLP dalam menganalisis ulasan produk. Analisis tersebut bertujuan untuk memberikan pemahaman yang lebih dalam tentang kinerja metode NLP dalam mengekstraksi informasi dari ulasan produk, serta membantu dalam memahami dampaknya terhadap pengambilan keputusan atau pemahaman tentang produk yang dianalisis.



**Gambar 1.** Alur Penelitian

Pada tahap penentuan produk, produk yang digunakan dalam penelitian ini adalah pakaian pria di Shopee , toko Forklass Classy Way’s Official Shop di jadikan sumber pengumpul data ulasan dengan hasil pencarian kata kunci “ Kaos *Oversize* Pria “. Produk dipilih berdasarkan urutan jumlah penjualan produk yang terdapat di toko Forklass Classy Way’s Official Shop. Tahapan pengumpulan data ulasan dilakukan dengan menggunakan Aplikasi Scraper yang dibuat secara mandiri menggunakan bahasa Python. Setiap produk membutuhkan 6 ulasan dari halaman pertama. Mulai dari review bintang satu hingga review bintang lima, sehingga menghasilkan total 3.500 review. Ulasan pembelian produk disimpan dalam basis data MySQL agar mudah dikelola [11]. Dari 3.500 ulasan yang didapat, dipilih ulasan dengan kriteria berikut:

1. Setiap produk diambil maksimal 1 ulasan pada setiap jumlah bintang yang memenuhi kriteria dan diurutkan berdasarkan ulasan terbaru.
2. Ulasan dengan panjang kurang dari 100 karakter tidak digunakan.
3. Ulasan yang digunakan hanyalah ulasan yang memiliki tag.
4. Ulasan yang hanya berisi emoticons atau simbol tidak digunakan.



- e. Stemming  
Stemming berfungsi untuk membuat suatu kata menjadi kata dasar, dengan menghilangkan semua imbuhan yang ada pada kata tersebut (Majid et al., 2023b). Sebagai contoh kalimat “bahannya halus dan tidak membuat kegerahan” kemudian diubah menjadi “bahan halus dan tidak buat gerah[14].
- f. Stopword Removal  
Stopword removal berfungsi untuk menghilangkan kata-kata yang mempunyai jumlah kemunculan banyak tapi tidak terlalu penting . Kata yang masuk stopwords 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”[15].

Dataset yang melalui langkah preprocessing dengan dan tanpa metode NLP dibagi terdiri dari 2 bagian dengan 208 data latih dan 52 data uji. Kemudian terapkan beban menggunakan algoritma Term Inverse Frekuensi (TF-IDF). TF-IDF digunakan untuk memberikan bobot hubungan antara kata atau istilah dan peringkat keseluruhan. KeteraturanMunculnya sebuah kata dalam sebuah ulasan menunjukkan pentingnya kata tersebut dalam ulasan dan Yang mana review mengandung kata-kata tersebut sehingga review dapat digolongkan menjadi 2 jenis (review ulasan positif dan negatif)[16]. Hitung TF-IDF menggunakan persamaan berikut:

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

Di mana  $W_{x,y}$  adalah bobot istilah (ty) terhadap dokumen (dx). Sedangkan  $tf_{x,y}$  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).

### 3. HASIL DAN PEMBAHASAN

#### 3.1 Daftar Ulasan Produk Shopee

Ulasan yang digunakan didapatkan dari kegiatan pencarian ulasan produk pada bulan oktober 2022. di Shopee. Ulasan diambil dari produk terlaris dengan pencarian menggunakan kata kunci “Kaos Oversize Pria”.Contoh ulasan yang digunakan ditunjukkan pada Tabel 1.

**Tabel 1.** Kuesioner Animasi Interaktif

No	Akun	Isi Ulasan	Tag Ulasan	Bintang	Label
1.	Diankun12	Proses Pengemasan dan pengiriman cukup cepat,packing aman,bahan kain adem,jahitan rapih pula.Aku beli ukuran XL untuk badan kecil ukuran tidak terlalu kebesaran banget. Kek lebih mirip Fuji kalo make baju oversize getoh.Thankyu sehat selalu	Kualitas Produk Sangat Baik	5/5	Positif
2.	C*****y	Bahan bagus banget kualitas produk nya sangat tidak mengecewakan pokok suka banget deh sama kaos ini dan Respond penjualan sangat baik banget pass aku salah pilih dan chat ke adminnya trus langsung di ganti dan langsung di kirim engga pake lama lagi udah nyampe aja barengnya	Kualitas Produk Sangat Baik	5/5	Positif
3.	anggiehanda yani1	Barang sesuai di tempat tujuan dengan aman. Menurut saya dengan harga segini okelah	Kualitas produk Baik	4/5	Positif

No	Akun	Isi Ulasan	Tag Ulasan	Bintang	Label
4.	sultanarrafa	yang size XL di aku bisa jadi dress Tag ga sesuai gambar di keterangan dan fototagnya h&m pas dating tagnya lain padahal buat ngado jadi gagal	Kualitas Produk Kurang Baik	2/5	Negatif
5.	S*****d	Warna printnya jelek, kaosnya bolong Size L sama XL kegedean kalo jualan gini trus gw yakin pembeli akan kapok	Kualitas Produk Kurang Baik	2/5	Negatif

Analisis Sentimen pada Ulasan Pembelian Produk di Marketplace Shopee Menggunakan Pendekatan *Natural Language Processing* Pada Tabel 1, akun Faris menuliskan ulasan tentang Ukuran pas, bahannya bagus yang ovz yang regular biasa, tetapi memberikan tag ulasan kualitas produk sangat baik dengan jumlah bintang 5/5 dan diberikan label positif. Akun achmad menuliskan ulasan tentang ukuran yang pas dan memberi tag kualitas produk baik dengan jumlah bintang 5/5 dan diberikan label negatif. Bisa disimpulkan bahwa jumlah bintang pada ulasan tidak selalu mewakili isi ulasan. Ulasan dengan jumlah bintang 5/5 tidak menjamin pembeli puas pada semua fitur barang dan pelayanan yang didapat. Fitur tag pada ulasan membantu pembeli untuk lebih spesifik menyebut fitur apa saja yang diulas. Informasi dari ulasan ini dapat diolah untuk menjadi perbaikan oleh penjual ke depan.

### 3.2 Analisis Ulasan

Opini yang digunakan sebanyak 200 opini yang diberikan oleh 5 orang reporter, dan dari opini tersebut diperoleh 117 opini positif dan 83 opini negatif. Evaluasi tersebut diubah menjadi dataset yang terdiri dari 152 data latih dan 48 data uji, kemudian dilakukan preprocessing dengan menggunakan metode NLP. Pada tahap preprocessing, fitur normalisasi kata, penghapusan kata dasar, dan penghapusan kata berhenti diterapkan pada setiap evaluasi. *Wordnormalizer* untuk menangani variasi ejaan kata yang mempunyai arti sama sehingga dihitung sebagai satu istilah [17]. 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 Tulisan	Word normalizer	Jumlah Ulasan
1.	Bgt	Banget	11
2.	Bgt	Banget	24
3.	Banget	Banget	5
4.	Gk	Tidak	14
5.	Ga	Tidak	10
6.	Tidak	Tidak	17
7.	G	Tidak	20
8.	Gak	Tidak	26
9.	Tidak	Tidak	9
10.	tdk	tidak	35

Sumber: Hasil Pengolahan Data (2022)

Tabel 2 menunjukkan variasi kata yang mempunyai makna yang sama. Kata "gak" ditemukan pada 26 ulasan dan kata "tdk" ditemukan pada 35 ulasan, digunakan pembeli untuk menggantikan kata "tidak". Kata "bgt" ditemukan pada 24 ulasan dan kata "banget" ditemukan pada 5 ulasan yang digunakan oleh pembeli untuk menggantikan kata "banget". Menggunakan penormal kata dapat menangani variasi kata yang banyak digunakan dalam teks ulasan pembeli untuk mengubahnya menjadi istilah yang sama.

Selanjutnya dilakukan proses stemming pada dataset tersebut, untuk menghilangkan awalan, sisipan, dan akhiran kata sehingga menjadi bentuk dasar, dengan tujuan agar pengambilan informasi menjadi lebih mudah dan efisien [17]. Contoh penerapan *stemming* di tunjukan pada tabel 3.

Tabel 3. Hasil penerapan *stemming*

No	Istilah Pada Ulasan	Stemming	Jumlah Ulasan
1	Pesan	Pesan	10
2	Memesan	Pesan	2
3	Dipesan	Pesan	5
4	Penjahit	Jahit	11
5	Dijahit	Jahit	4
6	Jahitannya	Jahit	1
7	Berlangganan	Langganan	2
8	Langganan	Langganan	27
9	Di Kirimkan	Kirim	9
10	Dikirim	Kirim	35

Pada Tabel 3, Anda dapat melihat kata-kata dalam ulasan seperti "dikirim" dan "dikirimkan".. Jika dihilangkan maka awalan, sisipan, dan akhiran menjadi kata dasar "dikirim".. Begitu pula dengan kata "dijahit", "penjahit" dan "jahitannya", setelah melalui proses pemotongan batang menjadi kata dasar "boleh".. Akar menyebabkan kata-kata dalam ulasan mengambil bentuk dasarnya dan menjadi istilah yang sama.. Proses penghapusan kata sandi kemudian dilakukan untuk menghapus kata sandi dari pertimbangan. ulasan maupun dalam jual beli secara daring. Contoh daftar stopwords yang dipakai ditunjukkan pada Tabel 4.

Tabel 4. Contoh Daftar Stopword

No	Stopword	Jumlah Kemunculan
1.	Ke	20
2.	Dengan	28
3.	Untuk	30
4.	Jadi	12
5.	Juga	50
6.	Ya	55
7.	Dan	70
8.	Tapi	67
9.	Di	88
10.	yang	94

Seperti terlihat pada Tabel 4, kata "yang" termasuk dalam daftar stopwords yang paling sering muncul sebanyak 94 kali, dan kata "di" sebanyak 88 kali.. Selain dalam bentuk kata, angka juga dimasukkan dalam stopwords, angka tersebut tidak berpengaruh dalam analisis sentimen dan dapat dihilangkan, untuk mengurangi noise dan meningkatkan 'efektivitas. Hasil keseluruhan pra-pemrosesan dapat dilihat pada Tabel 5.

Tabel 5. Hasil Pra-Pemrosesan

No	Ulasan Produk	Pra-Pemrosesan			Label Aktual
		Word Normalizer	Stemming	Stopword Removal	
1	Proses Pengemasan dan pengiriman cukup cepat,packing aman,bahan kain adem,jahitan rapih pula.Aku beli ukuran XL untuk badan kecil ukuran tidak terlalu kebesaran banget. Kek lebih mirip Fuji kalo make baju oversize getoh.Thankyu sehat selalu	Proses pengemasan dan pengiriman cukup cepat packing aman bahan kain adem jahitan rapih juga aku beli ukuran XL untuk badan kecil ukuran tidak terlalu kebesaran banget kaya lebih mirip fuji kalo makai baju oversize gitu thank you sehat selalu	Proses kemas dan kirim cukup cepat packing aman bahan kain adem jahit rapih juga aku beli ukur XL untuk badan kecil ukur tidak terlalu besar banget kaya lebih mirip fuji makai baju oversize gitu thank you sehat selalu	Cukup cepat packing aman bahan kain adem jahit rapih ukur tidak terlalu besar thank you sehat selalu	Positif

2	Bahan bagus banget kualitas produk nya sangat tidak mengecewakan pokok suka banget deh sama kaos ini dan Respond penjualan sangat baik banget pass aku salah pilih dan chat ke adminnya trus langsung di ganti dan langsung di kirim engga pake lama lagi udah nyampe aja barengnya	Bahan bagus banget kualitas produknya sangat tidak mengecewakan pokok suka banget sama kaos ini dan respoond penjual sangat baik banget pas aku salah pilih dan chat ke admin terus langsung di ganti dan langsung dikirim tidak pakai lama lagi udah sampai aja barangnya	Bahan bagus banget kualitas produk sangat tidak kecewa pokok suka banget sama kaos ini dan respond penjual sangat baik banget pas aku salah pilih dan chat admin terus langsung ganti tidak pakai lama udah sampai aja barang	Bahan bagus banget kualitas produk tidak kecewa suka banget sangat baik tidak pakai lama	Positif
3	Barang sesuai di tempat tujuan dengan aman. Menurut saya dengan harga segini okelah yang size XL di aku bisa jadi dress	Barang sesuai di tempat tujuan dengan aman menurut saya dengan harga segini okelah yang size XL di aku bisa jadi dress	Barang sesuai tempat tujuan dengan aman nurut saya harga gini oke size XL aku bisa jadi dress	Barang sesuai tempat tujuan aman harga oke	Positif
4	Tag ga sesuai gambar di keterangan dan fototagnya h&m pas dating tagnya lain padahal buat ngado jadi gagal	Tag tidak sesuai gambar di keterangan dan fotonya hm pas datang tagnya lain padahal buat kado jadi gagal	Tag tidak sesuai gambar keterangan dan foto hm pas datang tag lain padahal buat kado jadi gagal	Tidak sesuai gambar foto gagal	Negatif
5	Warna printnya jelek, kaosnya bolong Size L sama XL kegedean kalo jualan gini trus gw yakin pembeli akan kapok	Warna printnya jelek kaosnya bolong size L sama XL kegedean kalau jualan begini terus yakin pembeli akan kapok	Warna print jelek kaos bolong size L sama XL gede kalau jualan begini terus yakin beli akan kapok	Warna jelek jual begini kapok	Negatif

Pada Tabel 5 menunjukkan ulasan produk yang telah melewati pra-pemrosesan menggunakan pendekatan NLP dengan fitur word normalize, stemming, dan stopword removal. alah satu indikator kepuasan pelanggan adalah pencapaian performa dari sebuah produk yang diterima oleh pelanggan sama dengan ekspektasi pelanggan itu sendiri. Dilakukan pelanggan dengan memberikan ulasan menggunakan kata berkonotasi positif, hal ini sesuai dengan hasil ulasan yang diberi label oleh koresponden dimana ulasan terdapat kata “bagus”, “suka” dan “cepat”. Sedangkan indikator pelanggan tidak puas adalah dengan memberikan ulasan dengan kata berkonotasi negatif, seperti kata “tidak”, “kecewa”, dan “tipis”.

### 3.3 Perhitungan KNN Menggunakan TF-IDF

Penelitian ini menggunakan algoritma KNN dengan k=3 untuk mengklasifikasikan review.. Gunakan KNN yang didukung TF-IDF untuk mengukur bobot atau pentingnya sebuah kata di jurnal [18]. Perhitungan TF-IDF dilakukan terhadap 152 data latih dan 48 data uji pada status data setelah dilakukan preprocessing dengan dan tanpa metode NLP.. Kemudian dihitung jumlah kecocokan hasil pelabelan dari prediksi data uji 48 yang dibandingkan subjek dengan pelabelan manual menggunakan matriks konfusi .. Matriks konfusi membagi situasi menjadi: True Positive (TP) menunjukkan jumlah data dengan kelas positif dan hasil prediksi benar, True Negative (TN) menunjukkan banyaknya data dengan kelas negatif dan hasil prediksi benar, Positif palsu positif (FP) menunjukkan banyaknya data yang memberikan hasil positif.. kelas dan prediksi salah dan False Negatives (FN) menunjukkan banyaknya data dengan kelas negatif dan prediksi salah.. Hasil peramalan disajikan pada Tabel 6.

**Tabel 6.** Hasil confusion matrix data uji dengan dan tanpa pra-pemrosesan NLP

Hasil Klarifikasi	Tanpa Pra-pemrosesan	Dengan Pra-pemrosesan
True Positif	22	22
True Negatif	15	22
False Positif	5	5
False Negatif	10	8

Kemudian, hasil dari matriks konfusi dihitung sehingga diperoleh nilai presisi, presisi, dan recall. Akurasi adalah perbandingan sampel yang diklasifikasikan dengan benar dari dan jumlah total sampel. Presisi dihitung menggunakan Persamaan 2.. Presisi adalah rasio sampel positif yang diklasifikasikan dengan benar terhadap jumlah total sampel yang diharapkan positif, presisi dihitung menggunakan Persamaan 3.. Penarikan artikel adalah total sampel positif yang diklasifikasikan dengan benar sampel positif, recall dihitung menggunakan Persamaan 4[19].

$$Akurasi = \frac{TP+TN}{TP+FP+FN+TN} \tag{2}$$

$$Presisi = \frac{TP}{FP+TP} \tag{3}$$

$$Recall = \frac{TP}{TP+FN} \tag{4}$$

Persamaan 2, 3 dan 4 diterapkan pada hasil pengujian dua keadaan data uji, yaitu tanpa pendekatan NLP dan dengan pendekatan NLP, didapat hasil perbandingan yang ditunjukkan pada Tabel 7.

**Tabel 7.** Hasil pengujian dengan dan tanpa pra-pemrosesan NLP

Hasil Pengujian	Tanpa Pra-pemrosesan	Dengan Pra-pemrosesan
Akurasi	71,15%	77,19%
Presisi	81,48%	81,48%
Recal	68,75%	73,33%

Hasil percobaan pada Tabel 7 menunjukkan bahwa klasifikasi melalui preprocessing tanpa menggunakan metode NLP memberikan hasil akurasi sebesar 71,15%, sedangkan klasifikasi melalui preprocessing dengan metode NLP memberikan hasil akurasi sebesar 77,19%, akurasi meningkat sebesar 76,92%.. Akurasi kedua data tersebut memberikan hasil yang sama yaitu 81,48%.. Nilai recall data melalui preprocessing tanpa NLP menunjukkan hasil sebesar 68,75% dan melalui preprocessing dengan NLP menunjukkan peningkatan menjadi 73,33%.. Peningkatan hasil klasifikasi secara keseluruhan pada data yang telah mengalami preprocessing[20].

### 3.4 Frekuensi Kemunculan Kata

Kata-kata yang sering muncul dalam ulasan dapat menggambarkan penerimaan pasar secara keseluruhan terhadap produk tersebut. Kata-kata yang sering digunakan dalam ulasan Shopee untuk produk “kaos *oversize*” ditampilkan di cloud kata yang dihasilkan. dengan aplikasi Wordart pada Gambar 4 dan Gambar 5.



pembeli. Dengan memahami sentimen ini, pengecer dan penjual dapat merespons dengan lebih baik terhadap kebutuhan dan preferensi pelanggan, meningkatkan kualitas produk dan layanan mereka.

Nilai akurasi menjadi 77,19%, presisi 81,48%, dan *recall* 73,33%, hasil ini lebih tinggi di bandingkan dengan klasifikasi yang tidak menggunakan fitur NLP yang hanya menghasilkan nilai akurasi sebesar 71,15%, presisi 81,48%, dan *recall* 68,75%. Studi tersebut menemukan bahwa ulasan positif didominasi oleh kata “baik”, “suka”, “cepat”, “sangat”, “harga” dan “barang” menunjukkan bahwa produk kaos oversize mendapat ulasan *feedback* positif dari pembeli mengenai kualitas barang, kecepatan pengiriman dan harga barang. Sementara itu, beberapa kata Ulasan negatif didominasi oleh "tidak", "warna", "kirim", "kecewa" dan "berlangganan". menunjukkan bahwa produk kaos *oversize* mendapat *review* negatif dari pembeli karena ketidaksesuaian warna serta nomor atau data teknis barang yang dikirim bersamaan dengan barang yang dipesan pembeli. Dengan demikian, penelitian ini memberikan kontribusi penting dalam pemahaman tentang sentimen pengguna terhadap produk di marketplace Shopee dan memperkuat argumen untuk penggunaan NLP dalam analisis sentimen produk secara luas.

## REFERENCES

- [1] I. Setiawan, S. Nirwan, and F. Maisya Amelia, “Rancangan Bangun Aplikasi Marketplace Bagi Usaha Mikro Kecil, Dan Menengah Berbasis Web (SUB MODUL : PEMBELIAN),” 2018.
- [2] A. Alwendi, “Penerapan E-Commerce Dalam Meningkatkan Daya Saing Usaha,” *Manajemen Bisnis*, vol. 17, no. 3, pp. 317–325, 2020, [Online]. Available: <http://journal.undiknas.ac.id/index.php/magister-manajemen/>
- [3] Y. S. Mulyani, “E-Commerce Solusi Pemasaran UMKM Dalam Mengembangkan Industri Pariwisata Di Tengah Pandemi Covid-19 (Studi Kasus UMKM di Kota Tasikmalaya),” *Khasanah Ilmu - Jurnal Pariwisata Dan Budaya*, vol. 12, no. 2, pp. 131–141, 2021, doi: 10.31294/khi.v12i2.11293.
- [4] E. Y. Nasution, P. Hariani, L. S. Hasibuan, and W. Pradita, “Perkembangan Transaksi Bisnis E-Commerce terhadap Pertumbuhan Ekonomi di Indonesia,” *Jesya*, vol. 3, no. 2, pp. 506–519, 2020, doi: 10.36778/jesya.v3i2.227.
- [5] J. Sintadi, “Analisis Pengaruh Brand Image, Harga, Dan Review Produk Terhadap Keputusan Pembelian Smartphone Samsung Pada Mahasiswa Undip Semarang,” *Diponegoro Journal of Management*, vol. 8, no. 2, pp. 33–43, 2019, [Online]. Available: <http://ejournal-s1.undip.ac.id/index.php/djom>
- [6] N. Annisa, M. A. Nasution, and A. Firah, “Pengaruh Atribut Produk Dan Informasi Produk Terhadap Keputusan Pembelian Pada Pt . Trimitra Swadaya Medan the Effect of Product Attributes and Product Information on Purchase Decisions At Pt . Trimitra,” no. 1, pp. 67–76, 2023.
- [7] Z. Alhaq, A. Mustopa, S. Mulyatun, and J. D. Santoso, “Penerapan Metode Support Vector Machine Untuk Analisis Sentimen Pengguna Twitter,” *Journal of Information System Management (JOISM)*, vol. 3, no. 2, pp. 44–49, 2021, doi: 10.24076/joism.2021v3i2.558.
- [8] N. Munasatya and S. Novianto, “Natural Language Processing untuk Sentimen Analisis Presiden Jokowi Menggunakan Multi Layer Perceptron,” *Techno.Com*, vol. 19, no. 3, pp. 237–244, 2020, doi: 10.33633/tc.v19i3.3630.
- [9] R. Yusuf, T. A. Saputri, and A. A. Wicaksono, “Penerapan Natural Language Processing Berbasis Virtual Assistant Pada Bagian Administrasi Akademik Stmik Dharma Wacana,” *International Research on Big-Data and Computer Technology: I-Robot*, vol. 5, no. 1, pp. 33–47, 2022, doi: 10.53514/ir.v5i1.228.
- [10] V. Amrizal, “Penerapan Metode Term Frequency Inverse Document Frequency (Tf-Idf) Dan Cosine Similarity Pada Sistem Temu Kembali Informasi Untuk Mengetahui Syarah Hadits Berbasis Web (Studi Kasus: Hadits Shahih Bukhari-Muslim),” *Jurnal Teknik Informatika*, vol. 11, no. 2, pp. 149–164, 2018, doi: 10.15408/jti.v11i2.8623.
- [11] M. Ahmadar, P. Perwito, and C. Taufik, “Perancangan Sistem Informasi Penjualan Berbasis Web Rahayu Photo Copy Dengan Database MySQL,” *Dharmakarya*, vol. 10, no. 4, p. 284, 2021, doi: 10.24198/dharmakarya.v10i4.35873.
- [12] A. Majid, D. Nugraha, and F. D. Adhinata, “Sentiment Analysis on Tiktok Application Reviews Using Natural Language Processing Approach,” *Journal of Embedded Systems ...*, vol. 04, no. May, pp. 32–38, 2023, [Online]. Available: <https://journal.unm.ac.id/index.php/JESSI/article/view/471%0Ahttps://journal.unm.ac.id/index.php/JESSI/article/download/471/352>
- [13] E. Setiana, Marwondo, Venia Retreva Danestiara, and Wiyanudin, “Analisis Sentimen Pelaksanaan Kuliah Online

- Menggunakan Algoritma Support Vector Machine,” *Nuansa Informatika*, vol. 17, no. 2, pp. 66–70, 2023, doi: 10.25134/ilkom.v17i2.11.
- [14] H. R. Pramudita, “Penerapan Algoritma Stemming Nazief & Adriani Dan Similarity Pada Penerimaan Judul Thesis,” *Jurnal Ilmiah Data Manajemen dan Teknologi Informasi (DASI)*, vol. 15, no. 4, pp. 15–19, 2014.
- [15] R. Rinandyaswara, Y. A. Sari, and M. T. Furqon, “Pembentukan Daftar Stopword Menggunakan Term Based Random Sampling Pada Analisis Sentimen Dengan Metode Naïve Bayes (Studi Kasus: Kuliah Daring Di Masa Pandemi),” *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 9, no. 4, p. 717, 2022, doi: 10.25126/jtiik.2022934707.
- [16] R. Kosasih and A. Alberto, “Analisis Sentimen Produk Permainan Menggunakan Metode TF-IDF Dan Algoritma K-Nearest Neighbor,” *Infotekjar : Jurnal Nasional Informatika Dan Teknologi Jaringan*, vol. 6, no. 1, pp. 134–139, 2021, [Online]. Available: <https://doi.org/10.30743/infotekjar.v6i1.3893>
- [17] E. Utami, A. F. Iskandar, W. Hidayat, A. B. Prasetyo, and A. D. Hartanto, “Covid-19 Hoax Detection Using KNN in Jaccard Space,” *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, vol. 15, no. 3, p. 255, 2021, doi: 10.22146/ijccs.67392.
- [18] A. Zamsuri, E. Asril, M. Sadar, and Y. Turnandes, “Analisis sentimen opini vaksinasi Covid-19,” *ZONAsi: Jurnal Sistem Informasi*, vol. 5, no. 1, pp. 100–110, 2023.
- [19] A. Tharwat, “Classification Assessment Methods,” *Applied Computing and Informatics*, vol. 17, no. 1, pp. 168–192, 2018, doi: 10.1016/j.aci.2018.08.003.
- [20] M. K. Anam, B. N. Pikir, and M. B. Firdaus, “Penerapan Naïve Bayes Classifier, K-Nearest Neighbor (KNN) dan Decision Tree untuk Menganalisis Sentimen pada Interaksi Netizen dan Pemerintah,” *MATRIK : Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer*, vol. 21, no. 1, pp. 139–150, 2021, doi: 10.30812/matrik.v21i1.1092.