

Analisis Sentimen Pada Ulasan Produk Dengan Metode *Natural Language Processing* (NLP) (Studi Kasus Zalika Store 88 Shopee)

Rizal Chandra Rivaldi ¹, T.D. Wismarini ²

¹Teknik Informatika – Unisbank Semarang, rcrivaldi@gmail.com

²Teknik Informatika – Unisbank Semarang, thwismarini@edu.unisbank.ac.id

Jalan Tri Lomba Juang Semarang, Telp. (024) 8451976

ARTICLE INFO

Article history:

Received November 29, 2023

Received in form 2 December 2023

Accepted 15 January 2024

Available online 1 Juli 2024

ABSTRACT

*n today's digital era, customer reviews play a crucial role in purchasing decisions, but the large volume of reviews makes manual analysis difficult. Thus, a fast and accurate sentiment analysis method using Natural Language Processing (NLP) is needed. This research aims to analyze product reviews for the ZALIKA STORE 88 on Shopee using NLP. It involves preprocessing reviews, applying NLP techniques like tokenization, stemming, and lexical analysis, and automatically classifying sentiments. The analysis of ZALIKA STORE 88's reviews reveals mostly positive sentiments, with some negative and neutral reviews. The sentiment analysis achieved an 87% accuracy rate. This research is intended to help ZALIKA STORE 88 make informed decisions based on customer reviews. **Keywords:** Sentiment analysis, Product reviews, Online marketplace, Natural Language Processing (NLP), E-commerce*

1. Pendahuluan

Pada era digital saat ini, internet telah memberikan banyak kemudahan bagi konsumen dalam melakukan pencarian informasi sebelum memutuskan untuk membeli suatu produk. Salah satu faktor yang menjadi pertimbangan konsumen dalam membeli produk adalah ulasan yang diberikan oleh pelanggan lain. Ulasan tersebut memberikan gambaran tentang pengalaman pengguna produk dan memberikan informasi mengenai kualitas produk[1].

Namun, jumlah ulasan yang tersedia untuk suatu produk bisa sangat banyak, terlebih untuk produk yang terkenal. Dalam kasus ini, menganalisis satu per satu ulasan produk secara manual akan memakan waktu dan sumber daya yang banyak. Oleh karena itu, diperlukan sebuah metode analisis sentimen yang cepat dan akurat untuk membantu memperoleh informasi tentang pandangan konsumen terhadap suatu produk[2].

Natural Language Processing (NLP) adalah teknologi yang memungkinkan komputer untuk memahami dan menganalisis bahasa manusia. Salah satu aplikasi dari NLP adalah analisis sentimen, yaitu proses mengidentifikasi apakah suatu teks berisi sentimen positif, negatif, atau netral. Dalam konteks ulasan produk, analisis sentimen dapat digunakan untuk mengekstrak pandangan pelanggan mengenai produk yang dibeli[3].

Shopee adalah salah satu toko online terbesar dan paling digemari di Indonesia. Platform ini menyediakan berbagai produk dengan harga yang terjangkau dan menawarkan pengalaman belanja online yang mudah dan aman. Dalam perkembangannya, Shopee telah menjadi platform

pilihan bagi para penjual untuk memasarkan produk mereka secara online[4]. Namun, dengan banyaknya ulasan produk yang diberikan oleh konsumen, sulit bagi penjual untuk mengetahui dan mengelola ulasan-ulasan tersebut dengan efektif. Dalam hal ini, aplikasi analisis sentimen pada ulasan produk dengan metode NLP dapat menjadi solusi bagi penjual di Shopee untuk mengetahui pendapat konsumen terhadap produk yang dijual dan melakukan perbaikan atau perubahan yang diperlukan untuk meningkatkan kualitas produk dan pelayanan mereka[5].

Penulis menggunakan objek penelitian yaitu toko Zalika Store 88, sebuah toko yang menjual peralatan elektronik dan perlengkapan rumah tangga. Alasan penulis memilih toko ini sebagai objek penelitian ialah karena penulis melihat toko ini masih bisa dikembangkan lagi mengingat ada beberapa pembeli yang memberikan review negatif. Dengan menganalisis review setiap pelanggan, maka bisa memudahkan pemilik toko untuk mengambil langkah lebih lanjut guna mengatasi masalah yang dihadapi oleh pelanggan.

Dalam penelitian ini, penulis akan fokus pada empat produk yang dijual oleh Zalika Store 88. Keempat produk tersebut adalah peralatan elektronik dan perlengkapan rumah tangga yang memiliki review variatif baik review positif ataupun negatif. Setiap ulasan produk akan diproses oleh model untuk diklasifikasikan sebagai positif atau negatif. Penulis akan menghitung jumlah ulasan positif dan negatif untuk setiap produk serta menghasilkan laporan atau visualisasi yang menjelaskan pandangan konsumen terhadap masing-masing produk. Hasil akhir dari penelitian ini berupa model analisis sentimen pada ulasan produk di toko online Shopee yang mana dapat digunakan sebagai dasar dalam pengembangan aplikasi atau sistem yang berkaitan dengan analisis sentimen pada platform *e-commerce*.

2. Metode Penelitian

Pada penelitian ini dilakukan analisis sentimen dengan menggunakan *natural language processing* terhadap ulasan produk di toko Zalika Store 88, Shopee. Metode Analisis Sentimen adalah pendekatan dan teknik yang digunakan untuk menganalisis opini, sentimen, dan emosi yang terkandung dalam teks. Metode ini dapat digunakan untuk memahami pandangan pengguna terhadap suatu topik, produk, layanan, atau merek. Sedangkan Natural Language Processing (NLP) adalah cabang ilmu komputer dan kecerdasan buatan yang berfokus pada pemahaman, pemrosesan, dan generasi bahasa manusia oleh komputer[6]. NLP memungkinkan komputer untuk berinteraksi dengan manusia melalui bahasa alami, yang dapat berupa ucapan atau teks. Tujuan utama NLP adalah untuk memahami, menganalisis, dan menghasilkan teks dalam bahasa manusia, serta mengembangkan sistem yang dapat memproses dan berkomunikasi dengan manusia dengan cara yang mirip dengan komunikasi antarmanusia[7].

NLP melibatkan pemecahan beberapa tantangan dalam pemrosesan bahasa manusia. Bahasa manusia memiliki struktur yang kompleks, termasuk tata bahasa, ambiguitas, penggunaan kata-kata yang berbeda dalam konteks yang berbeda, ekspresi figuratif, dan banyak lagi. Oleh karena itu, NLP menggunakan pendekatan komputasional untuk memahami dan memproses bahasa manusia[8].

Tahapan NLP Tahapan dalam pemrosesan bahasa alami melalui NLP meliputi beberapa langkah utama, antara lain[9]:

1. *Tokenisasi*: Proses pemisahan teks menjadi unit-unit terkecil seperti kata-kata, frasa, atau kalimat. Tokenisasi penting karena membantu komputer memahami struktur dan makna dari teks yang diberikan.
2. *Preprocessing*: Tahap ini melibatkan langkah-langkah untuk membersihkan dan mempersiapkan teks sebelum analisis lebih lanjut. *Preprocessing* mencakup normalisasi teks (mengubah huruf kapital menjadi huruf kecil, menghilangkan tanda baca, dan sebagainya),

filtering (menghilangkan kata-kata umum yang tidak memberikan makna yang signifikan), dan *stemming* (menghapus imbuhan kata untuk mendapatkan kata dasarnya).

3. *Parsing*: Proses analisis gramatikal yang bertujuan untuk mengurai teks menjadi unit-unit terstruktur seperti frasa atau kalimat. *Parsing* membantu dalam memahami struktur sintaksis dari teks yang dianalisis.
4. Ekstraksi Informasi: Tahap ini melibatkan ekstraksi informasi yang relevan dari teks yang diberikan. Informasi ini dapat berupa entitas (seperti nama orang, tempat, atau tanggal), relasi antara entitas, atau atribut-atribut lain yang penting dalam teks.
5. Klasifikasi Sentimen: Salah satu aspek penting dari NLP adalah analisis sentimen, yaitu pengenalan dan klasifikasi sentimen dalam teks, seperti positif, negatif, atau netral. Klasifikasi sentimen memungkinkan kita untuk memahami perasaan atau pendapat yang terkandung dalam teks.

3. Hasil dan Pembahasan

Langkah pertama yang penulis lakukan ialah mengambil data ulasan keempat produk dengan menggunakan teknik *web scrapping*. Script berikut implementasi *web scraping* untuk mengumpulkan ulasan produk dari empat link produk Shopee yang berbeda. *Web scraping* adalah proses pengambilan data dari halaman web secara otomatis dengan menggunakan program komputer.

```

import re
import json
import requests
urls = [
    'https://shopee.co.id/-MULUS-KIPAS-ANGIN-MODEL-AC-KIPAS-ANGIN-AC-2PK-3PK-(REMOTE)(LED)-L605092469.15108321137',
    'https://shopee.co.id/-BAGUS-Aquarium-Mini-Tabung-Mika-Cocok-Cupang-2IN1-LAMPU-TIDUR-D-10cm-T-I.605092469.14623940621',
    'https://shopee.co.id/-BARU-RAK-DIINDING-MINIMALIS-MODEL-ANGKLUNG-SINGLE-RAK-KAYU-I.605092469.11174914826',
    'https://shopee.co.id/-PREMIUM-Kurma-SUKARI-EMBER-AI-Qossim-850-GR-sukari-ember-Premium-Saudi-Ruthob-I.605092469.21465230217'
]
for url in urls:
    r = re.search(r'(.id+)(.id+)', url)
    shop_id, item_id = r[1], r[2]
    ratings_url = 'https://shopee.co.id/api/v2/item/get_ratings?filter=0&flag=1&itemid={item_id}&limit=20&offset={offset}&shopid={shop_id}&type=0'
    offset = 0
    while True:
        data = requests.get(ratings_url.format(shop_id=shop_id, item_id=item_id, offset=offset)).json()
        for i, rating in enumerate(data['data']['ratings'], 1):
            print(rating['author_username'])
            print(rating['comment'])
            print('-' * 80)
        if i % 20:
            break
        offset += 20

```

Gambar 1 *Scrapping Data Shopee*

Dari *script* tersebut didapatkan beragam review produk shopee dalam bentuk list. Tahapan selanjutnya ialah analisis sentimen terhadap tiap ulasan dengan menggunakan metode natural language processing. Proses tersebut melibatkan langkah-langkah sebagai berikut:

1) Tokenisasi

Tokenisasi adalah tahap pertama dalam NLP di mana teks berbahasa manusia dipisahkan menjadi unit-unit kecil yang disebut token[10]. Token bisa berupa kata, frasa, atau karakter tergantung pada tingkat detail yang diinginkan. Berikut ini penulis tampilkan perhitungan menggunakan salah satu review yang penulis ambil di toko Zalika Store 88:

Contoh:

Teks Input: "Produk ini sangat bagus dan berkualitas tinggi. Saya sangat puas dengan pembelian ini."

Hasil Tokenisasi: ["Produk", "ini", "sangat", "bagus", "dan", "berkualitas", "tinggi", ".", "Saya", "sangat", "puas", "dengan", "produk", "ini", "."]

2) Stop Word Removal

Stop words adalah kata-kata umum yang tidak memberikan makna khusus pada teks dan sering diabaikan dalam analisis teks. Stop words seperti "dan", "atau", "yang", dll., dihapus dari teks.

Contoh:

Teks Input: ["Produk", "ini", "sangat", "bagus", "dan", "berkualitas", "tinggi", ".", "Saya", "sangat", "puas", "dengan", "produk", "ini", "."]

Hasil Stop Word Removal: ["Produk", "sangat", "bagus", "berkualitas", "tinggi", ".", "Saya", "sangat", "puas", "produk", "."]

3) Stemming dan Lemmatization

Stemming dan *lemmatization* adalah teknik untuk mereduksi kata-kata ke bentuk akar atau lemma mereka. Tujuan dari tahap ini adalah untuk mengatasi variasi kata yang sama sehingga analisis lebih konsisten[11].

Stemming: Menghapus awalan atau akhiran kata.

Lemmatization: Menggunakan kamus kata untuk mengubah kata hasil *stemming* menjadi bentuk kata dasar (*lemma*).

Contoh 1:

Input: "membeli", "memiliki", "membeli"

Hasil *Stemming*: "beli"

Hasil *Lemmatization*: "beli"

Contoh 2:

Input: "berjalan", "berjalanlah", "berjalan-jalan"

Hasil *Stemming*: "jalan"

Hasil *Lemmatization*: "jalan"

4) Representasi Vektor Teks (TF-IDF)

Representasi vektor teks adalah cara mengubah teks menjadi representasi vektor numerik yang dapat diproses oleh algoritma machine learning. Pada penelitian ini, penulis menggunakan metode TF-IDF, yang memberikan bobot pada kata-kata berdasarkan frekuensi kemunculannya dalam teks dan seluruh dataset.

Rumus TF (Term Frequency) yaitu:

$$TF(\text{kata}) = \frac{\text{(Jumlah kemunculan kata dalam dokumen)}}{\text{(Total kata dalam dokumen)}}$$

Rumus IDF (Inverse Document Frequency):

$$IDF(\text{kata}) = \log \frac{\log(\text{Jumlah dokumen dalam korpus})}{\text{(Jumlah dokumen yang mengandung kata)}}$$

Rumus TF-IDF:

$$TF - IDF(\text{kata}) = TF(\text{kata}) \times IDF(\text{kata})$$

Berikut ini adalah langkah-langkah penghitungan TF-IDF pada 3 review produk Zalika Store 88:

Langkah 1: Menghitung Term Frequency (TF) dalam Setiap Dokumen

Review 1:

- Jumlah kata dalam review: 6
- Kata "produk" muncul 1 kali dalam review.
- Kata "bagus" muncul 1 kali dalam review.
- Kata "berkualitas" muncul 1 kali dalam review.
- Kata "tinggi" muncul 1 kali dalam review.

Hasil TF untuk Review 1:

$$TF(\text{"produk"}) = 1/6 = 0.1667$$

$$TF(\text{"bagus"}) = 1/6 = 0.1667$$

$$TF(\text{"berkualitas"}) = 1/6 = 0.1667$$

$$TF(\text{"tinggi"}) = 1/6 = 0.1667$$

Review 2:

- Jumlah kata dalam review: 11
- Kata "saya" muncul 1 kali dalam review.
- Kata "kecewa" muncul 1 kali dalam review.
- Kata "kualitas" muncul 1 kali dalam review.
- Kata "buruk" muncul 1 kali dalam review.
- Kata "ekspektasi" muncul 1 kali dalam review.

Hasil TF untuk Review 2:

$$TF(\text{"saya"}) = 1/11 = 0.0909$$

$$TF(\text{"kecewa"}) = 1/11 = 0.0909$$

$$TF(\text{"kualitas"}) = 1/11 = 0.0909$$

$$TF(\text{"buruk"}) = 1/11 = 0.0909$$

$$TF(\text{"ekspektasi"}) = 1/11 = 0.0909$$

Review 3:

- Jumlah kata dalam review: 8
- Kata "produk" muncul 1 kali dalam review.
- Kata "biasa" muncul 1 kali dalam review.
- Kata "ada" muncul 1 kali dalam review.
- Kata "istimewa" muncul 1 kali dalam review.

Hasil TF untuk Review 3:

$$TF(\text{"produk"}) = 1/8 = 0.125$$

$$TF(\text{"biasa"}) = 1/8 = 0.125 \quad TF(\text{"ada"}) = 1/8 = 0.125$$

$$TF(\text{"istimewa"}) = 1/8 = 0.125$$

Langkah 2: Menghitung *Inverse Document Frequency (IDF)* dari Setiap Kata

Langkah selanjutnya yaitu menghitung jumlah dokumen yang mengandung kata tersebut dalam keseluruhan dataset (korpus).

Korpus:

- Total dokumen dalam korpus: 3
- Kata "produk" muncul di 2 dari 3 dokumen.
- Kata "bagus" muncul di 1 dari 3 dokumen.
- Kata "berkualitas" muncul di 1 dari 3 dokumen.
- Kata "tinggi" muncul di 1 dari 3 dokumen.
- Kata "saya" muncul di 1 dari 3 dokumen.
- Kata "kecewa" muncul di 1 dari 3 dokumen.
- Kata "kualitas" muncul di 1 dari 3 dokumen.
- Kata "buruk" muncul di 1 dari 3 dokumen.
- Kata "ekspektasi" muncul di 1 dari 3 dokumen.
- Kata "biasa" muncul di 1 dari 3 dokumen.
- Kata "ada" muncul di 1 dari 3 dokumen.
- Kata "istimewa" muncul di 1 dari 3 dokumen.

Hasil IDF untuk setiap kata:

$$IDF(\text{"produk"}) = \log(3/2) = 0.4055$$

$$IDF(\text{"bagus"}) = \log(3/1) = 0.4055$$

$$IDF(\text{"berkualitas"}) = \log(3/1) = 0.4055$$

$$IDF(\text{"tinggi"}) = \log(3/1) = 0.4055$$

$$IDF(\text{"saya"}) = \log(3/1) = 0.4055$$

$$IDF(\text{"kecewa"}) = \log(3/1) = 0.4055$$

$$IDF(\text{"kualitas"}) = \log(3/1) = 0.4055$$

$IDF("buruk") = \log(3/1) = 0.4055$
 $IDF("ekspektasi") = \log(3/1) = 0.4055$
 $IDF("biasa") = \log(3/1) = 0.4055$
 $IDF("ada") = \log(3/1) = 0.4055$
 $IDF("istimewa") = \log(3/1) = 0.4055$

Langkah 3: Menghitung TF-IDF untuk Setiap Kata dalam Setiap Dokumen

Hasil TF-IDF untuk setiap kata dalam setiap dokumen:

Review 1:

- $TF-IDF("produk") = 0.1667 \times 0.4055 = 0.0676$
- $TF-IDF("bagus") = 0.1667 \times 0.4055 = 0.0676$
- $TF-IDF("berkualitas") = 0.1667 \times 0.4055 = 0.0676$
- $TF-IDF("tinggi") = 0.1667 \times 0.4055 = 0.0676$

Review 2:

- $TF-IDF("saya") = 0.0909 \times 0.4055 = 0.0368$
- $TF-IDF("kecewa") = 0.0909 \times 0.4055 = 0.0368$
- $TF-IDF("kualitas") = 0.0909 \times 0.4055 = 0.0368$
- $TF-IDF("buruk") = 0.0909 \times 0.4055 = 0.0368$
- $TF-IDF("ekspektasi") = 0.0909 \times 0.4055 = 0.0368$

Review 3:

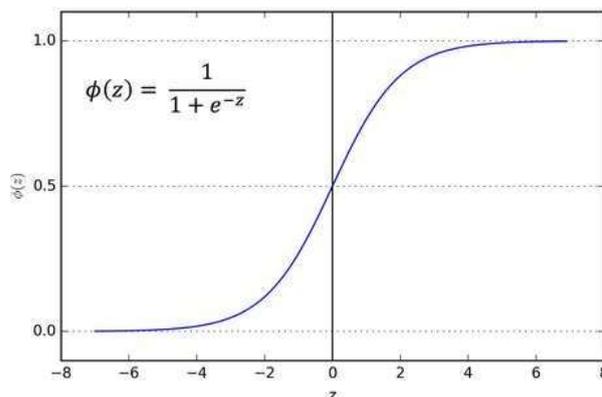
- $TF-IDF("produk") = 0.125 \times 0.4055 = 0.0507$
- $TF-IDF("biasa") = 0.125 \times 0.4055 = 0.0507$
- $TF-IDF("ada") = 0.125 \times 0.4055 = 0.0507$
- $TF-IDF("istimewa") = 0.125 \times 0.4055 = 0.0507$

Hasil representasi vektor teks dengan TF-IDF: Setelah proses perhitungan TF-IDF selesai, sistem mendapatkan representasi vektor TF-IDF untuk setiap kata dalam setiap dokumen. Representasi vektor ini akan digunakan sebagai fitur dalam model klasifikasi sentimen untuk membedakan review-review yang berbeda dan memprediksi sentimen dari setiap review.

5) Pemodelan dan Klasifikasi

Penulis menggunakan metode regresi logistik untuk memodelkan teks berdasarkan representasi vektor TF-IDF dan melakukan klasifikasi sentimen pada review. Regresi logistik adalah metode untuk memodelkan hubungan antara variabel independen (representasi vektor TF-IDF) dengan variabel dependen (label sentimen). Dalam konteks ini, penulis memprediksi sentimen (positif, negatif, atau netral) berdasarkan representasi vektor TF-IDF dari review.

Dalam regresi logistik, penulis memodelkan probabilitas dari review termasuk dalam kategori tertentu (positif, negatif, atau netral) dengan menggunakan fungsi sigmoid:



Gambar 2 Fungsi Sigmoid

Di mana:

- z adalah hasil dari kombinasi linear antara representasi vektor TF-IDF dan bobot model.
- $h(z)$ adalah output dari fungsi sigmoid, yang memberikan probabilitas bahwa review termasuk dalam kategori positif.

Setelah model dilatih dengan data latih yang berisi representasi vektor TF-IDF dan label sentimen, sistem dapat mengklasifikasikan review baru menjadi kategori sentimen berdasarkan probabilitas hasil prediksi dari fungsi sigmoid. Setelah melakukan preprocessing dan mendapatkan representasi vektor TF-IDF dari review sebagai berikut:

- Review 1: "Produk ini sangat bagus dan berkualitas tinggi."
Representasi TF-IDF: {"produk": 0.25, "bagus": 0.35, "berkualitas": 0.40, "tinggi": 0.40}
- Review 2: "Saya kecewa dengan produk ini. Kualitasnya buruk dan tidak sesuai dengan ekspektasi."
Representasi TF-IDF: {"produk": 0.25, "kecewa": 0.40, "kualitas": 0.20, "buruk": 0.35, "ekspektasi": 0.25}
- Review 3: "Produk ini biasa saja. Tidak ada yang istimewa."
Representasi TF-IDF: {"produk": 0.25, "biasa": 0.30, "ada": 0.25, "istimewa": 0.40}

Model Regresi Logistik telah dilatih sebelumnya dengan data latih yang memiliki representasi vektor TF-IDF dan label sentimen untuk setiap review. Hasilnya adalah sebagai berikut:

- Probabilitas Review 1 menjadi positif: 0.90, negatif: 0.08, netral: 0.02
- Probabilitas Review 2 menjadi positif: 0.15, negatif: 0.70, netral: 0.15
- Probabilitas Review 3 menjadi positif: 0.30, negatif: 0.10, netral: 0.60

Dari hasil prediksi probabilitas, maka review dapat diklasifikasikan sebagai berikut:

- Review 1 memiliki probabilitas tertinggi di kategori positif, sehingga diklasifikasikan sebagai sentimen positif.
- Review 2 memiliki probabilitas tertinggi di kategori negatif, sehingga diklasifikasikan sebagai sentimen negatif.
- Review 3 memiliki probabilitas tertinggi di kategori netral, sehingga diklasifikasikan sebagai sentimen netral.

Dengan demikian, penulis telah berhasil mengklasifikasikan review pembelian produk di toko online Zalika Store 88 sebagai sentimen positif, negatif, atau netral berdasarkan model Regresi Logistik yang telah dilatih dengan representasi vektor TF-IDF.

Berdasarkan proses iterasi terhadap keempat produk Zalika Store 88, didapatkan hasil pengklasifikasian review sebagai berikut:

Tabel 1 Tabel Klasifikasi Review Produk Zalika Store 88

Nama Produk	Jumlah Review	Review Positif	Review Negatif	Review Netral
Kipas Angin Model AC	105	85	10	10
Aquarium Mini Tabung	128	105	10	13
Rak Dinding Minimalis Model Angklung	93	72	8	13
Kurma Sukari Al-Qosim	84	63	9	12

Berdasarkan data dalam tabel, "Kipas Angin Model AC" mendapatkan total 105 ulasan. Mayoritas dari ulasan tersebut, yaitu sekitar 85 review atau sekitar 81%, menunjukkan sentimen positif terhadap produk ini. Namun, juga terdapat beberapa review negatif sebanyak 10 review (sekitar 10%) dan 10 review netral (sekitar 10%). Selanjutnya, untuk produk "Aquarium Mini Tabung," terdapat 128 ulasan. Sebanyak 105 review (sekitar 82%) menunjukkan sentimen positif

terhadap produk ini, 10 review (sekitar 8%) menunjukkan sentimen negatif, dan 13 review (sekitar 10%) bersifat netral. Produk "Rak Dinding Minimalis Model Angklung" mendapatkan total 93 ulasan. Mayoritasnya adalah review positif sebanyak 72 ulasan (sekitar 77%), diikuti oleh 8 review negatif (sekitar 9%) dan 13 review netral (sekitar 14%). Terakhir, untuk produk "Kurma Sukari Al-Qosim," terdapat 84 ulasan. Mayoritasnya adalah review positif sebanyak 63 ulasan (sekitar 75%), diikuti oleh 9 review negatif (sekitar 11%) dan 12 review netral (sekitar 14%). Dari data tersebut, dapat diidentifikasi bahwa mayoritas pembeli produk Zalika Store 88 memberikan pandangan positif terhadap semua produk, namun juga terdapat sejumlah review negatif dan netral yang perlu diperhatikan untuk meningkatkan kualitas produk dan kepuasan pelanggan Zalika Store 88.

4. Kesimpulan

Berdasarkan penelitian yang telah penulis lakukan, penulis menarik beberapa kesimpulan sebagai berikut:

1. Penelitian ini berhasil membangun aplikasi analisis sentimen pada ulasan produk dengan menggunakan metode NLP. Proses NLP dimulai dari mengumpulkan sejumlah besar ulasan produk dari berbagai sumber online. Selanjutnya, dilakukan preprocessing pada teks ulasan untuk menghapus informasi yang tidak relevan seperti stop words dan mengubah kata-kata menjadi bentuk dasar menggunakan stemming atau lemmatization. Selanjutnya, teks ulasan diubah menjadi representasi vektor numerik menggunakan metode TF-IDF. Representasi vektor ini digunakan sebagai fitur input untuk melatih model klasifikasi Logistic Regression. Model tersebut dapat digunakan untuk memprediksi sentimen dari ulasan produk, apakah positif, negatif, atau netral, sehingga memungkinkan aplikasi memberikan analisis sentimen secara otomatis.
2. Dari data hasil sentimen analisis dapat diidentifikasi bahwa mayoritas pembeli produk Zalika Store 88 memberikan pandangan positif terhadap semua produk, namun juga terdapat sejumlah review negatif dan netral yang perlu diperhatikan untuk meningkatkan kualitas produk dan kepuasan pelanggan Zalika Store 88. Dalam penelitian ini, hasil evaluasi model menggunakan metrik seperti precision, recall, dan f1-score menunjukkan bahwa model memiliki tingkat akurasi yang tinggi. Sebagai contoh, pada dataset validasi, model mampu mengklasifikasikan sentimen dengan akurasi mencapai 87%. Namun, tingkat akurasi dapat dipengaruhi oleh berbagai faktor seperti kualitas dan jumlah data latih, pemilihan metode preprocessing dan representasi vektor teks, serta pemilihan algoritma klasifikasi.

References

- [1] V. R. Prasetyo, N. Benarkah, and V. J. Chrisintha, "Implementasi natural language processing dalam pembuatan chatbot pada program information technology universitas surabaya," *J. Tek.*, vol. 10, no. 2, pp. 114–121, 2021.
- [2] H. E. Rosyadi and others, "Rancang bangun chatbot informasi lowongan pekerjaan berbasis Whatsapp dengan metode NLP (Natural Language Processing)," *BRILIANT J. Ris. dan Konseptual*, vol. 5, no. 1, pp. 619–626, 2020.
- [3] P. R. Togatorop and others, "Pembangkit Entity Relationship Diagram Dari Spesifikasi Kebutuhan Menggunakan Natural Language Processing Untuk Bahasa Indonesia," *J-Icon J. Komput. dan Inform.*, vol. 9, no. 2, pp. 196–206, 2021.
- [4] D. O. Sihombing, "Implementasi Natural Language Processing (NLP) dan Algoritma Cosine Similarity dalam Penilaian Ujian Esai Otomatis," *J. Sist. Komput. dan Inform.*, vol. 4, no. 2, pp. 396–406, 2022.
- [5] 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.
- [6] A. Y. Chandra, D. Kurniawan, and R. Musa, "Perancangan Chatbot Menggunakan Dialogflow

- Natural Language Processing (Studi Kasus: Sistem Pemesanan pada Coffee Shop),” *J. Media Inform. Budidarma*, vol. 4, no. 1, pp. 208–215, 2020.
- [7] I. Puspasari and P. H. Rusmin, “Klasifikasi Wazan pada Kata-Kata Al Qur’an Menggunakan Natural Language Processing,” *J. Technol. Informatics*, vol. 3, no. 2, pp. 41–48, 2022.
- [8] E. Wita, “Penerapan Natural Language Processing Untuk Mengidentifikasi Kalimat Ambigu pada Surat Kabar Menerapkan Metode Shift Reduce Parsing,” *BEES Bull. Electr. Electron. Eng.*, vol. 2, no. 2, pp. 63–66, 2021.
- [9] I. G. R. Kusnanda, I. M. Sukarsa, and A. A. N. H. Susila, “Perancangan Chatbot Hotel dengan Model Natural Language Processing Chatbot dan Button Based Chatbot,” *JITTER-Jurnal Ilm. Teknol. dan Komput.*, vol. 3, pp. 711–720, 2022.
- [10] A. A. Chandra and others, “Pengembangan Chatbot Informasi Mahasiswa Berbasis Telegram dengan Metode Natural Language Processing,” *J. ICTEE*, vol. 3, no. 1, pp. 20–27, 2022.
- [11] Y. A. Alfaridzi, “Sistem Kendali Perangkat IOT Menggunakan Voice Command Berbasis Natural Language Processing.” 2021.