

# Analisis Sentimen Ulasan Produk Tokopedia Menggunakan IndoBERT Laporan Final Project - Big Data dan AI

Muhammad Iqbal Arrasyid

1304221038

[ibalarrasyid@student.telkomuniversity.ac.id](mailto:ibalarrasyid@student.telkomuniversity.ac.id)

June 6, 2025



PROGRAM STUDI S1 INFORMATIKA PJJ  
FAKULTAS INFORMATIKA  
UNIVERSITAS TELKOM  
BANDUNG  
2025

# Contents

<b>1</b>	<b>Deskripsi Kasus</b>	<b>3</b>
<b>2</b>	<b>Tujuan dan Manfaat</b>	<b>3</b>
<b>3</b>	<b>Metodologi</b>	<b>4</b>
3.1	Teknologi dan Tools . . . . .	4
3.2	Model . . . . .	5
3.3	Langkah-langkah Analisis . . . . .	6
<b>4</b>	<b>Dokumentasi Kode</b>	<b>7</b>
4.1	Inisialisasi dan Import Library . . . . .	7
4.2	Memuat dan Menampilkan Data . . . . .	7
4.3	Pra-pemrosesan Teks . . . . .	7
4.4	Analisis Sentimen Menggunakan IndoBERT . . . . .	8
4.5	Visualisasi WordCloud dan Sentimen . . . . .	8
<b>5</b>	<b>Masalah dan Solusi</b>	<b>9</b>
<b>6</b>	<b>Hasil Analisis</b>	<b>10</b>
<b>7</b>	<b>Kesimpulan</b>	<b>13</b>
<b>8</b>	<b>Pengembangan Selanjutnya</b>	<b>13</b>

# 1 Deskripsi Kasus

Tokopedia merupakan salah satu platform e-commerce terbesar di Indonesia yang secara aktif menyediakan fitur ulasan produk bagi seluruh penggunanya. Melalui fitur ini, para pembeli dapat menuliskan pengalaman mereka setelah melakukan transaksi, baik terkait kualitas produk, kecepatan pengiriman, maupun pelayanan penjual. Setiap ulasan yang ditulis oleh pengguna tidak hanya memberikan masukan yang sangat berarti bagi calon pembeli lainnya, tetapi juga membantu penjual dalam mengevaluasi dan meningkatkan kualitas layanan serta produk yang mereka tawarkan. Dengan membaca ulasan-ulasan tersebut, calon pembeli dapat mempertimbangkan keputusan pembelian secara lebih matang, sedangkan penjual dapat memahami harapan dan kebutuhan pelanggan secara lebih mendalam.

Dalam upaya untuk memahami pola kepuasan pelanggan dan mengidentifikasi area yang perlu diperbaiki, analisis sentimen terhadap ulasan-ulasan ini menjadi sangat penting. Melalui analisis sentimen, kita dapat mengelompokkan opini pelanggan ke dalam kategori positif, netral, atau negatif, sehingga perusahaan dapat mengambil langkah strategis berdasarkan data yang objektif. Selain itu, analisis ini juga dapat membantu Tokopedia untuk memantau reputasi produk dan penjual di platform mereka secara lebih efisien.

Untuk mendukung penelitian ini, penulis menggunakan dataset yang bersumber dari [Kaggle](#) dengan judul *Tokopedia Product Reviews*. Dataset ini berisi ribuan ulasan pelanggan yang dikumpulkan dari berbagai transaksi di Tokopedia. Setiap entri dalam dataset tersebut memuat informasi berupa teks ulasan, rating yang diberikan oleh pembeli, nama produk yang diulas, serta kategori produk terkait. Dengan memanfaatkan dataset ini, penulis dapat melakukan analisis secara komprehensif terhadap berbagai aspek yang memengaruhi kepuasan pelanggan di Tokopedia. Selain itu, penulis juga dapat mengidentifikasi tren sentimen yang berkembang di antara pengguna, sehingga hasil analisis ini dapat memberikan wawasan yang bermanfaat bagi pengembangan fitur dan peningkatan layanan di platform Tokopedia.

## 2 Tujuan dan Manfaat

Penelitian ini memiliki beberapa tujuan utama yang saling berkaitan dan memberikan manfaat strategis bagi berbagai pihak yang terlibat di ekosistem Tokopedia. Pertama, penulis ingin mendeteksi opini pelanggan secara otomatis melalui pemanfaatan teknologi analisis sentimen berbasis model IndoBERT. Dengan mengotomatiskan proses identifikasi sentimen dari ribuan ulasan yang masuk setiap harinya, penulis dapat mempercepat dan mempermudah pemetaan opini pelanggan terhadap produk maupun layanan yang tersedia di platform tersebut. Selain itu, otomatisasi ini juga memungkinkan perusahaan untuk memantau perubahan persepsi pelanggan secara real-time.

Selanjutnya, penulis berusaha memberikan insight yang relevan dan dapat diandalkan untuk mendukung pengambilan keputusan bisnis. Melalui analisis mendalam terhadap data ulasan, penulis dapat mengungkap pola-pola sentimen yang muncul di berbagai kategori produk. Dengan demikian, manajemen Tokopedia dan para penjual dapat merumuskan strategi pemasaran, perbaikan layanan, serta pengembangan produk baru berdasarkan data yang valid dan terukur. Selain itu, insight yang dihasilkan dari penelitian ini dapat membantu perusahaan dalam mengidentifikasi masalah yang sering dikeluhkan pelanggan, sehingga mereka dapat mengambil tindakan korektif secara lebih cepat dan tepat sasaran.

Terakhir, penulis ingin mengukur tingkat kepuasan pelanggan secara umum berdasarkan ulasan yang telah dikumpulkan. Dengan melakukan analisis kuantitatif terhadap distribusi sentimen positif, netral, dan negatif, penulis dapat memperoleh gambaran menyeluruh mengenai persepsi pelanggan terhadap kualitas layanan Tokopedia. Selain itu, hasil pengukuran ini juga dapat digunakan sebagai tolok ukur keberhasilan berbagai inisiatif yang telah dijalankan oleh perusahaan dalam rangka meningkatkan pengalaman pengguna. Melalui tujuan-tujuan tersebut, penulis berharap penelitian ini dapat memberikan kontribusi nyata dalam pengembangan ekosistem e-commerce di Indonesia, khususnya dalam hal peningkatan kualitas layanan dan kepuasan pelanggan.

## 3 Metodologi

### 3.1 Teknologi dan Tools

Dalam menjalankan proyek analisis sentimen ulasan produk Tokopedia, penulis memilih beberapa teknologi dan tools yang saling melengkapi untuk memastikan proses berjalan secara efektif dan efisien. Pada tahap awal, penulis menggunakan **Apache Spark** sebagai platform utama untuk pemrosesan data dalam skala besar. Spark penulis pilih karena kemampuannya memproses data secara paralel dan terdistribusi, sehingga dapat menangani ribuan hingga jutaan ulasan produk dengan cepat. Pengalaman penulis menunjukkan bahwa tanpa Spark, pengolahan data besar akan sangat lambat dan sulit dioptimalkan. Dengan Spark, penulis dapat membagi beban komputasi ke beberapa node, sehingga proses pembersihan data, transformasi, hingga inferensi model berjalan lebih lancar dan skalabel.

Selanjutnya, penulis menggunakan **Python** sebagai bahasa pemrograman utama dalam proyek ini. penulis memilih Python karena fleksibilitasnya yang tinggi dan dukungan ekosistem pustaka yang sangat lengkap, terutama untuk pengolahan data dan machine learning. Dengan Python, penulis dapat menulis kode dengan cepat dan mudah dipahami, serta memanfaatkan library populer seperti pandas untuk manipulasi data, PySpark untuk integrasi dengan Spark, dan transformers untuk pemodelan NLP. Pengalaman penulis dalam menggunakan Python mempercepat proses pengembangan dan eksperimen, sehingga penulis bisa fokus pada analisis dan interpretasi hasil.

Kemudian, penulis mengintegrasikan **HuggingFace Transformers** untuk mengakses model IndoBERT yang telah dilatih khusus untuk Bahasa Indonesia. Library ini menyediakan antarmuka yang intuitif dan fleksibel untuk melakukan pemodelan Natural Language Processing, termasuk klasifikasi sentimen. Dengan HuggingFace, penulis dapat dengan mudah memanggil model IndoBERT, melakukan inferensi, dan mengelola pipeline analisis tanpa perlu membangun model dari awal. Hal ini sangat membantu dalam mempercepat proses pengembangan dan menjaga kualitas hasil analisis sentimen yang relevan dengan konteks lokal.

Terakhir, untuk mendukung eksperimen dan komputasi yang membutuhkan sumber daya tinggi, penulis menggunakan **Google Colab** sebagai lingkungan komputasi berbasis cloud. Google Colab menyediakan akses gratis ke GPU, yang penulis manfaatkan untuk mempercepat proses inferensi model IndoBERT yang cukup berat secara komputasi. Selain itu, Google Colab memungkinkan penulis menjalankan kode secara interaktif dan berbagi hasil eksperimen dengan mudah. Dengan kombinasi teknologi dan tools ini, penulis berhasil membangun pipeline analisis sentimen yang tidak hanya efektif dan efisien, tetapi juga fleksibel dan mudah dikembangkan lebih lanjut sesuai kebutuhan

proyek.

## 3.2 Model

Dalam proyek ini, penulis menggunakan model berikut:

`w11wo/indonesian-roberta-base-sentiment-classifier`

Model ini merupakan bagian dari keluarga IndoBERT dan dirancang khusus untuk tugas klasifikasi sentimen dalam bahasa Indonesia. Sebelum membahas lebih lanjut, penting untuk memahami apa itu IndoBERT dan arsitektur RoBERTa Base yang menjadi fondasi model ini.

**IndoBERT** adalah model bahasa alami (Natural Language Processing) yang dikembangkan khusus untuk bahasa Indonesia. Secara teknis, IndoBERT merupakan singkatan dari Indonesia Bidirectional Encoder Representations from Transformers. Model ini dibangun menggunakan framework deep learning PyTorch dan mengadopsi arsitektur Transformer yang memungkinkan pemahaman konteks kata secara dua arah, yaitu membaca konteks dari kiri dan kanan secara simultan. Keunggulan utama IndoBERT terletak pada kemampuannya memahami struktur, pola, dan nuansa bahasa Indonesia yang kompleks, termasuk variasi dialek dan morfologi yang khas. Model ini dilatih menggunakan korpus besar berbahasa Indonesia yang beragam, seperti Wikipedia Indonesia, artikel berita, dan data web lokal, sehingga mampu menangkap konteks dan makna kata dengan lebih akurat dibandingkan model multibahasa umum seperti Multilingual BERT (M-BERT). Berkat pendekatan pre-training dan fine-tuning yang tepat, IndoBERT telah terbukti unggul dalam berbagai tugas NLP berbahasa Indonesia, termasuk analisis sentimen, deteksi hoaks, dan pengenalan emosi.

Sementara itu, **RoBERTa Base** adalah salah satu varian model Transformer yang dikembangkan sebagai penyempurnaan dari BERT (Bidirectional Encoder Representations from Transformers). RoBERTa menghilangkan beberapa batasan pada proses pre-training BERT dengan menggunakan dataset yang lebih besar dan teknik pelatihan yang lebih agresif, seperti penghapusan tugas Next Sentence Prediction dan penggunaan batch yang lebih besar. Arsitektur RoBERTa Base terdiri dari 12 lapisan encoder Transformer, dengan 768 dimensi representasi dan 12 attention heads. Model ini dikenal karena kemampuannya dalam memahami konteks kalimat secara mendalam dan menghasilkan representasi teks yang kaya secara semantik. Dengan kata lain, RoBERTa Base dapat menangkap hubungan antar kata dan frasa dalam sebuah kalimat dengan lebih baik, sehingga meningkatkan performa di berbagai tugas NLP.

Model `w11wo/indonesian-roberta-base-sentiment-classifier` menggabungkan kekuatan IndoBERT dan arsitektur RoBERTa Base. Setelah menjalani proses pre-training pada data besar berbahasa Indonesia, model ini di-fine-tune menggunakan dataset *SmSA* yang berisi ribuan komentar dan ulasan dalam bahasa Indonesia.

Proses fine-tuning ini berlangsung selama lima epoch, di mana model secara bertahap belajar mengenali pola-pola sentimen positif, netral, dan negatif dengan lebih presisi. Selama pelatihan, penulis mengamati penurunan loss yang konsisten dan kestabilan pada data validasi, yang menandakan bahwa model berhasil menggeneralisasi dengan baik tanpa mengalami overfitting. Hasil evaluasi menunjukkan bahwa model mencapai akurasi sebesar 94,36% dan skor F1-makro sebesar 92,42%, angka yang menunjukkan performa sangat baik dalam klasifikasi sentimen. Angka ini juga mengindikasikan bahwa model

mampu memberikan prediksi yang seimbang dan akurat di ketiga kelas sentimen, sehingga dapat diandalkan untuk analisis ulasan produk Tokopedia.

Lebih jauh lagi, penulis memanfaatkan library HuggingFace Transformers untuk mengakses dan mengimplementasikan model ini dalam pipeline analisis sentimen. Library ini menyediakan antarmuka yang intuitif dan fleksibel, memungkinkan penulis untuk melakukan inferensi secara efisien dan mengintegrasikan model dengan framework lain tanpa kesulitan. Dengan menggunakan model ini, penulis dapat secara otomatis mengklasifikasikan ribuan ulasan produk ke dalam kategori sentimen yang relevan, sehingga memperoleh wawasan yang akurat mengenai persepsi pelanggan terhadap produk dan layanan di Tokopedia.

Pengalaman menggunakan model ini juga mengajarkan penulis pentingnya pemilihan model yang sesuai konteks bahasa dan domain. IndoBERT, sebagai model yang sudah dioptimasi untuk bahasa Indonesia, mampu menangkap makna tersirat dan variasi ekspresi yang khas dalam ulasan pelanggan. Hal ini berbeda jauh dengan model generik yang sering kali gagal memahami konteks lokal dan menghasilkan prediksi yang kurang tepat. Oleh karena itu, penggunaan IndoBERT memberikan keunggulan kompetitif dalam analisis sentimen berbahasa Indonesia, sekaligus menjadi fondasi yang kuat untuk pengambilan keputusan bisnis yang lebih tepat dan berbasis data.

Secara keseluruhan, model `indoBert` tidak hanya membantu penulis dalam mengolah data ulasan dengan akurat, tetapi juga membuka peluang untuk pengembangan lebih lanjut, seperti fine-tuning pada domain produk tertentu atau integrasi dengan sistem monitoring real-time. Dengan demikian, model ini menjadi salah satu komponen kunci dalam membangun solusi analisis sentimen yang efektif dan relevan untuk kebutuhan e-commerce di Indonesia.

### 3.3 Langkah-langkah Analisis

1. **Inisialisasi Spark:** Membuat session Spark.
2. **Memuat Dataset:** Dataset ulasan dimuat ke dalam Spark DataFrame.
3. **Pra-pemrosesan Teks:**
  - Mengubah semua huruf menjadi huruf kecil.
  - Menghapus simbol dan angka.
  - Membersihkan spasi berlebih.
4. **Analisis Sentimen:**
  - Menggunakan model `w11wo/indonesian-roberta-base-sentiment-classifier`.
  - Menerapkan model melalui `pandas_udf` di Spark.
5. **Visualisasi Hasil:**
  - Wordcloud untuk frekuensi kata.
  - Diagram pie untuk distribusi sentimen.

## 4 Dokumentasi Kode

### 4.1 Inisialisasi dan Import Library

Python Code

```
1 from pyspark.sql import SparkSession
2 from pyspark.sql.functions import col, udf
3 from pyspark.sql.types import StringType
4 import re
5
6 spark = SparkSession.builder \
7     .appName("TokopediaSentimentAnalysis") \
8     .config("spark.driver.memory", "4g") \
9     .getOrCreate()
```

### 4.2 Memuat dan Menampilkan Data

Python Code

```
1 df = spark.read.option("header", "true").csv("product_reviews_dirty
    .csv")
2 df.show(5)
```

### 4.3 Pra-pemrosesan Teks

Python Code

```
1 def clean_text(text):
2     if text:
3         text = text.lower()
4         text = re.sub(r"[^a-zA-Z0-9\s]", "", text)
5         text = re.sub(r"\s+", " ", text).strip()
6         return text
7     return ""
8
9 clean_udf = udf(clean_text, StringType())
10 df_cleaned = df.withColumn("cleaned_text", clean_udf(col("text")))
```

## 4.4 Analisis Sentimen Menggunakan IndoBERT

### Python Code

```
1 from transformers import pipeline
2 import pandas as pd
3 from pyspark.sql.functions import pandas_udf
4
5 def get_sentiment_indobert(texts: pd.Series) -> pd.Series:
6     classifier = pipeline("sentiment-analysis",
7                             model="w11wo/indonesian-roberta-base-sentiment-classifier")
8
9     def _analyze(text):
10         if text:
11             try:
12                 result = classifier(text[:512])[0]
13                 return result['label'].lower()
14             except:
15                 return "neutral"
16         return "neutral"
17
18     return texts.apply(_analyze)
19
20 @pandas_udf(StringType())
21 def sentiment_udf(texts: pd.Series) -> pd.Series:
22     return get_sentiment_indobert(texts)
23
24 df_final = df_cleaned.withColumn("sentiment", sentiment_udf(col
    ("cleaned_text")))
```

## 4.5 Visualisasi WordCloud dan Sentimen

### Python Code

```
1 from wordcloud import WordCloud
2 import matplotlib.pyplot as plt
3
4 all_text = ' '.join([t["text"] for t in df_final.select("text")
5                     .filter(col("text").isNotNull()).collect()])
6
7 wordcloud = WordCloud(width=800, height=400, background_color='
    white').generate(all_text)
8
9 plt.figure(figsize=(12,6))
10 plt.imshow(wordcloud, interpolation='bilinear')
11 plt.axis("off")
12 plt.title("WordCloud - Tokopedia Review Produk")
13 plt.show()
```



#### Python Code

```
1 sentiment_counts = df_final.groupby("sentiment").count().toPandas()
2
3 plt.figure(figsize=(6, 6))
4 plt.pie(sentiment_counts['count'],
5         labels=sentiment_counts['sentiment'],
6         autopct='%1.1f%%',
7         startangle=140,
8         colors=['#66b3ff', '#99ff99', '#ff9999'])
9 plt.title('Distribusi Sentimen Ulasan Tokopedia')
10 plt.axis('equal')
11 plt.show()
```

## 5 Masalah dan Solusi

Selama proses analisis sentimen terhadap ulasan produk Tokopedia, penulis menghadapi beberapa tantangan teknis yang cukup signifikan. Namun, dengan pendekatan yang tepat dan solusi yang terstruktur, penulis berhasil mengatasi hambatan tersebut sehingga proses analisis dapat berjalan dengan lancar dan menghasilkan output yang akurat.

Pertama, penulis menemukan bahwa beberapa ulasan dalam dataset ternyata kosong atau memiliki panjang teks yang sangat pendek. Kondisi ini menyebabkan model IndoBERT mengalami error saat melakukan prediksi, karena model membutuhkan konteks yang cukup untuk mengidentifikasi sentimen dengan baik. Menyadari hal ini, penulis segera menambahkan mekanisme validasi panjang teks sebelum memasukkan data ke dalam pipeline prediksi. Dengan cara ini, penulis dapat memfilter dan mengabaikan ulasan yang tidak memenuhi ambang batas minimal panjang karakter atau kata. Selain itu, untuk ulasan yang terlalu singkat namun masih dianggap relevan, penulis melakukan penanganan khusus agar model tetap dapat memberikan prediksi yang wajar. Langkah ini terbukti efektif dalam mengurangi error dan menjaga kualitas hasil analisis.

Selanjutnya, penulis menghadapi kendala performa ketika memproses dataset yang berisi ratusan ribu baris menggunakan Apache Spark, terutama karena model IndoBERT termasuk model yang cukup berat dan membutuhkan sumber daya komputasi yang besar. Proses inferensi menjadi lambat dan berpotensi menyebabkan bottleneck dalam pipeline analisis. Untuk mengatasi masalah ini, penulis menerapkan dua strategi utama. Pertama, penulis melakukan sampling data untuk menguji dan mengoptimalkan pipeline sebelum menjalankan analisis pada keseluruhan dataset. Kedua, penulis menggunakan teknik batch prediksi, yaitu memproses data dalam kelompok-kelompok kecil secara paralel, sehingga mengurangi beban komputasi sekaligus mempercepat waktu eksekusi. Dengan kombinasi kedua metode ini, penulis berhasil meningkatkan efisiensi pemrosesan data tanpa mengorbankan akurasi hasil analisis.

Terakhir, saat melakukan visualisasi menggunakan wordcloud untuk menampilkan kata-kata yang paling sering muncul dalam ulasan, penulis mendapati bahwa hasilnya memunculkan banyak kata-kata umum dan tidak relevan yang mengganggu interpretasi. Kata-kata seperti “dan”, “yang”, atau “sangat” mendominasi visualisasi sehingga mengaburkan insight yang ingin diperoleh. Untuk mengatasi hal ini, penulis melakukan penyaringan kata dengan menyesuaikan daftar stopwords khusus Bahasa Indonesia. penulis

juga menambahkan filter untuk menghilangkan kata-kata umum yang tidak memberikan makna signifikan dalam konteks analisis sentimen. Proses ini melibatkan iterasi dan pengujian berulang agar wordcloud yang dihasilkan benar-benar merepresentasikan kata-kata kunci yang relevan dan informatif. Berkat penyesuaian ini, visualisasi menjadi lebih bersih dan mudah dipahami, sehingga membantu dalam mengidentifikasi tema dan sentimen utama dari ulasan pelanggan.

Secara keseluruhan, pengalaman mengatasi berbagai masalah teknis tersebut memperkuat pemahaman penulis tentang tantangan nyata dalam penerapan analisis sentimen pada data besar berbahasa Indonesia. Dengan pendekatan yang sistematis dan solusi yang tepat, penulis berhasil memastikan bahwa proses analisis berjalan optimal dan menghasilkan insight yang bermanfaat bagi pengembangan layanan Tokopedia.

## 6 Hasil Analisis

Dalam proses analisis sentimen terhadap ulasan produk Tokopedia, penulis menemukan pola distribusi sentimen yang cukup menarik dan memberikan gambaran jelas mengenai persepsi pelanggan terhadap produk yang mereka beli. Dari keseluruhan ulasan yang dianalisis, sekitar **79,8%** menunjukkan sentimen positif. Angka ini mengindikasikan bahwa mayoritas pelanggan merasa puas dengan produk dan layanan yang mereka terima. Mereka sering kali mengekspresikan rasa senang dan rekomendasi yang positif, yang tentunya menjadi sinyal baik bagi penjual dan platform Tokopedia dalam mempertahankan kualitas layanan mereka.

Namun, tidak semua ulasan mencerminkan kepuasan. Sekitar **11,3%** ulasan mengandung sentimen negatif, yang umumnya terkait dengan masalah pengiriman yang terlambat atau kualitas barang yang tidak sesuai dengan ekspektasi pelanggan. Melalui analisis ini, penulis dapat mengidentifikasi isu-isu spesifik yang sering menjadi sumber ketidakpuasan, sehingga penjual dapat fokus melakukan perbaikan di area tersebut. Misalnya, beberapa pelanggan mengeluhkan kemasan yang kurang aman atau barang yang rusak saat diterima, yang kemudian dapat menjadi perhatian utama dalam meningkatkan layanan logistik dan kualitas produk.

Selain itu, sekitar **8,9%** ulasan tergolong netral, biasanya berupa deskripsi produk tanpa menyertakan opini atau perasaan pelanggan. Ulasan netral ini sering kali berisi informasi faktual seperti spesifikasi produk atau pengalaman penggunaan yang tidak menimbulkan kesan positif maupun negatif. Meskipun demikian, ulasan netral tetap memberikan nilai karena membantu calon pembeli mendapatkan gambaran objektif mengenai produk.

Untuk memperkuat pemahaman terhadap data ini, penulis menyertakan visualisasi dalam bentuk wordcloud dan grafik distribusi sentimen. Pada gambar pertama (Word-Cloud - Tokopedia Review Produk), wordcloud menampilkan kata-kata yang paling sering muncul dalam ulasan pelanggan secara keseluruhan. Terlihat bahwa kata-kata seperti “barang”, “sesuai”, “terima kasih”, “cepat”, dan “pengiriman” mendominasi. Hal ini mencerminkan bahwa sebagian besar ulasan menekankan aspek positif seperti kecepatan pengiriman, kesesuaian produk dengan deskripsi, dan rasa puas terhadap layanan.

Selanjutnya, pada gambar kedua (Distribusi Sentimen Ulasan Tokopedia), grafik pie chart memperlihatkan proporsi sentimen dalam ulasan produk. Terlihat bahwa mayoritas ulasan bersentimen positif (79.8%), diikuti oleh sentimen negatif (11.3%) dan netral (8.9%). Visualisasi ini memberikan gambaran yang jelas bahwa sebagian besar pelanggan

merasa puas terhadap pengalaman belanja mereka di Tokopedia.

Dengan menggabungkan visualisasi ini, penulis dapat menyampaikan cerita yang koheren dan mudah dipahami mengenai persepsi pelanggan. Hal ini membantu pemangku kepentingan, seperti penjual dan manajemen platform, untuk lebih memahami kekuatan layanan yang sudah baik dan mengenali peluang perbaikan.

Secara keseluruhan, pendekatan ini memastikan bahwa informasi dari data ulasan dapat diakses dan dimengerti oleh berbagai pihak, termasuk mereka yang tidak memiliki latar belakang teknis. Dengan begitu, analisis ini dapat menjadi dasar yang kuat dalam pengambilan keputusan strategis berbasis data.

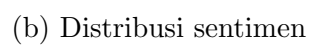


Figure 1: Visualisasi Sentimen Ulasan Produk Tokopedia

12

## 7 Kesimpulan

Setelah melalui serangkaian proses analisis yang mendalam, penulis menemukan bahwa model IndoBERT benar-benar menunjukkan efektivitas yang luar biasa dalam mengklasifikasikan sentimen dari ulasan produk yang ada di Tokopedia. Pada awal penelitian, penulis menyadari bahwa tantangan utama terletak pada kompleksitas bahasa Indonesia yang kaya akan variasi dialek, kosakata, dan ekspresi yang unik. Namun, dengan menggunakan model IndoBERT yang memang sudah dioptimasi khusus untuk bahasa Indonesia, penulis berhasil mengatasi hambatan tersebut dan mendapatkan hasil klasifikasi yang akurat serta relevan.

Dalam praktiknya, model ini mampu menangkap opini pelanggan dengan sangat baik, baik yang bersifat positif, netral, maupun negatif. Hal ini memungkinkan penjual dan platform Tokopedia untuk memperoleh insight yang sangat penting mengenai tingkat kepuasan pelanggan. Misalnya, penulis dapat mengidentifikasi produk-produk yang mendapat banyak ulasan negatif karena masalah kualitas atau pengiriman, sehingga penjual dapat segera melakukan perbaikan. Di sisi lain, produk dengan sentimen positif yang tinggi dapat menjadi bahan promosi yang efektif. Dengan demikian, model IndoBERT tidak hanya berfungsi sebagai alat analisis, tetapi juga menjadi pendorong pengambilan keputusan bisnis yang lebih cerdas dan berbasis data.

Selain itu, penulis menyadari bahwa kekuatan utama dari model ini terletak pada kemampuannya memahami konteks dan nuansa bahasa Indonesia secara mendalam. Berbeda dengan metode tradisional yang sering kali hanya mengandalkan kata kunci atau aturan sederhana, IndoBERT mampu menangkap makna tersirat dan ekspresi yang lebih kompleks dalam ulasan pelanggan. Hal ini memberikan nilai tambah yang signifikan dalam menghasilkan analisis sentimen yang lebih akurat dan bermakna.

## 8 Pengembangan Selanjutnya

Melihat hasil yang menjanjikan tersebut, penulis percaya bahwa masih banyak peluang untuk mengembangkan sistem analisis sentimen ini agar lebih bermanfaat dan aplikatif di masa depan. Pertama-tama, penulis merekomendasikan penerapan model ini pada data real-time. Bayangkan jika Tokopedia dapat memantau sentimen pelanggan secara langsung saat ulasan baru masuk. Dengan begitu, penjual dan tim manajemen dapat segera merespons perubahan opini pelanggan, baik dengan memperbaiki masalah yang muncul ataupun memanfaatkan feedback positif untuk meningkatkan strategi pemasaran. Hal ini tentunya akan meningkatkan kecepatan dan ketepatan dalam pengambilan keputusan, sehingga pengalaman pelanggan menjadi lebih baik.

Selanjutnya, penulis melihat potensi besar dalam mengintegrasikan hasil analisis sentimen ke dalam dashboard analitik berbasis web. Dengan adanya dashboard interaktif, berbagai pemangku kepentingan seperti manajer produk, tim pemasaran, dan penjual dapat dengan mudah mengakses data sentimen, melihat tren yang berkembang, dan menggali insight secara visual. Dashboard ini akan menjadi alat yang powerful untuk mendukung pengambilan keputusan yang cepat dan tepat, serta memudahkan komunikasi antar tim dalam mengelola produk dan layanan.

Terakhir, penulis menyarankan untuk melakukan fine-tuning model IndoBERT berdasarkan domain produk tertentu. Misalnya, model dapat disesuaikan secara khusus untuk kategori elektronik, fashion, atau makanan. Dengan pendekatan ini, model akan lebih peka terhadap istilah-istilah dan konteks yang khas pada masing-masing domain produk. Se-

bagai contoh, kata-kata yang memiliki makna positif dalam konteks elektronik mungkin berbeda dengan yang ada di kategori makanan. Dengan fine-tuning yang tepat, akurasi klasifikasi sentimen dapat meningkat secara signifikan, sehingga analisis menjadi lebih spesifik dan bernilai bagi penjual maupun platform.

Secara keseluruhan, pengembangan-pengembangan tersebut akan memperkuat kemampuan sistem analisis sentimen dan memberikan nilai tambah yang lebih besar dalam memahami kebutuhan serta preferensi pelanggan secara lebih detail dan kontekstual. penulis optimis bahwa dengan terus mengembangkan teknologi ini, Tokopedia dapat semakin meningkatkan kualitas layanan dan memperkuat posisi sebagai salah satu platform e-commerce terdepan di Indonesia.