e-ISSN : 2597-3673 (Online) , p-ISSN : 2579-5201 (Printed)
Vol.9 No.1 (June 2025)
Journal of Information System, Informatics and Computing
Website/URL: http://journal.stmikjayakarta.ac.id/index.php/jisicom
Email: jisicom@stmikjayakarta.ac.id , jisicom2017@gmail.com

PENERAPAN HADOOP DALAM ANALISIS SENTIMEN ULASAN PENGGUNA DI PLATFROM ECCOMERCE (*Application Of Hadoop In Sentiment Analysis Of User Reviews On The Eccomerce Platfrom*)

Nurdian Kasim [1], Ni Luh Ica Ardini[2], Alfi Zahrah Muharramah[3], Hikma[4], Muhammad Vannes Al Qadri[5], Rosalina[6], Wa Ode Asriyani[7], Eviriawan[8], Adha Mashur Sajiah[9]

Program Studi Teknik Informatika, Universitas Halu Oleo, Kendari
Jl.H.E.A Mokodompit Kampus Baru Tridharma Anduonohu,Kendari 92131 Indonesia

nurdiankasim1948@gmail.com, niluhicaardini@gmail.com, alfizahramuharramah@gmail.com, hikmaaaa36@gmail.com, muhammadvannesalqadri@gmail.com, rosalina29042004@gmail.com, waodeasriyani.e1e122034@gmail.com, eviriawan052004@gmail.com, adha.m.sajiah@uho.ac.id

Received: February 12, 2025**. Revised:** March 3, 2025**. Accepted:** March 5, 2025**. Issue Period:** Vol.9 No.1 (2025), Pp.11-24

Abstrak: Studi ini menyelidiki penggunaan teknologi Hadoop dan algoritma Naive Bayes untuk menganalisis sentimen ulasan pengguna di platform e-commerce. Data yang digunakan berasal dari 391.500 ulasan dari aplikasi Shopee yang dikumpulkan melalui scraping Google Play Store. Implementasi model klasifikasi sentimen, pengumpulan data melalui web scraping, dan pra-pemrosesan data menggunakan PySpark adalah metodologi penelitian. Hasil penelitian menunjukkan bahwa model Naive Bayes dapat mengklasifikasikan perasaan pengguna dengan akurasi 87%. Menurut analisis word cloud, elemen seperti gratis ongkir dan kemudahan penggunaan menjadi pendorong utama sentimen positif. Sementara itu, sentimen negatif didominasi oleh masalah teknis aplikasi dan layanan pelanggan. Penelitian ini menunjukkan bahwa penggunaan Hadoop dan Naive Bayes dalam analisis data ulasan berskala besar saat mengembangkan platform *e-commerce* adalah efektif.

Kata kunci: Analisis Sentimen, , Big Data, *E-commerce*, Hadoop, Naive Bayes, PySpark

Abstract: *This study examines the use of Hadoop technology and the Naive Bayes algorithm to analyze user sentiment on e-commerce platforms. The data used comes from 391.500 ulasan from the Shopee application, which was obtained via scraping the Google Play Store. The implementation of the classification model, data collection by web scraping, and data preparation using PySpark are research methodologies. The study's findings indicate that the Naive Bayes model can classify user behavior with an accuracy of 87%. According to word cloud analysis, features*

DOI: 10.52362/jisicom.v9i1.1784

e-ISSN : 2597-3673 (Online) , p-ISSN : 2579-5201 (Printed)
Vol.9 No.1 (June 2025)
Journal of Information System, Informatics and Computing
Website/URL: http://journal.stmikjayakarta.ac.id/index.php/jisicom
Email: jisicom@stmikjayakarta.ac.id , jisicom2017@gmail.com

like free ongkir and ease of use are the most positive indicators. On the other hand, negative sentiment was caused by issues with application technology and customer service. This study shows that using Hadoop and Naive Bayes to analyze large-scale data while developing an e-commerce platform is effective.
Keywords: *Big Data, E-commerce, Hadoop, Naive Bayes, PySpark, Sentiment Analysis*

I. PENDAHULUAN

Perilaku masyarakat telah berubah secara signifikan sebagai akibat dari meningkatnya pengguna internet dan perkembangan teknologi digital, khususnya di bidang ekonomi. Saat ini, salah satu media utama untuk belanja online adalah platform dan pasar e-commerce. Dengan sekitar 21 juta orang terlibat dalam transaksi e-commerce, Indonesia mengalami peningkatan transaksi pembelian online sebesar 14,9% pada tahun 2022, menunjukkan perkembangan penting dalam adopsi teknologi digital secara masyarakat di negara ini [1]. Pertumbuhan ekonomi yang berkelanjutan di tingkat nasional dan internasional juga sangat dipengaruhi oleh teknologi digital. Hal ini konsisten dengan penelitian Badan Pusat Statistik, yang mengklaim bahwa penggunaan teknologi informasi dan internet secara signifikan meningkatkan penjualan produk dan efisiensi perusahaan [2].

E-commerce telah berkembang pesat, tetapi juga memunculkan kesulitan baru yang harus diselesaikan, terutama di bidang pengalaman pelanggan [3]. Ketidakcocokan antara deskripsi produk dan kondisi aktual dari barang yang diperoleh, kualitas produk yang buruk, pembatasan pembayaran, dan masalah pengiriman adalah beberapa masalah utama yang sering ditemui pelanggan di platform e-commerce. Pelanggan sering menyuarakan semua masalah ini dalam evaluasi yang diposting di situs web e-commerce. Mengingat kuantitas evaluasi yang terus meningkat, sangat penting bagi bisnis untuk memahami bagaimana perasaan pelanggan tentang barang atau jasa yang mereka sediakan [4].

Analisis sentimen, yang mencoba mengklasifikasikan teks yang disediakan pengguna ke dalam tiga kategori utama positif, negatif, atau netral sangat penting untuk memahami opini konsumen. Bisnis dapat meningkatkan barang, layanan, dan pengalaman pengguna mereka dengan mengambil langkah yang tepat berdasarkan pemahaman mereka tentang sentimen pelanggan. Analisis sentimen juga berguna untuk merencanakan dan melihat tren pasar. [5]

Namun, teknologi yang dapat menangani data dalam jumlah besar juga disebut sebagai Big Data diperlukan untuk melakukan analisis sentimen pada ulasan pengguna ini. Kumpulan data yang sangat besar, rumit, dan tidak terstruktur yang tidak dapat dipahami oleh sistem atau aplikasi database tradisional disebut sebagai "Big Data" [6]. Akibatnya, kerangka kerja open source seperti Hadoop sangat penting untuk menangani dan menganalisis data dalam jumlah besar. Dengan kapasitas ini, Hadoop dapat mempercepat analisis sentimen evaluasi pengguna, memungkinkan bisnis untuk membuat keputusan berdasarkan data lebih cepat [7].

Selain itu, telah ditunjukkan bahwa teknik pembelajaran mesin seperti Naive Bayes berhasil mengkategorikan sentimen ulasan pengguna. Teorema Bayes, yang merupakan dasar dari algoritma probabilistik Naive Bayes, membuat asumsi bahwa setiap fitur data adalah unik [8]. Naive Bayes tetap dapat menghasilkan hasil yang luar biasa dalam klasifikasi teks, terutama dalam analisis sentimen, meskipun asumsi ini tidak umum dalam kehidupan nyata. Menurut sebuah studi oleh [9], Naive Bayes masih dapat menghasilkan hasil yang akurat ketika berhadapan dengan berbagai macam input teks. Naive Bayes, di sisi lain, sangat sukses dalam mengevaluasi ulasan produk di platform e-commerce Indonesia dan dapat membedakan dengan tepat antara sentimen positif dan negatif, menurut penelitian oleh [10].

Terlepas dari akurasi tinggi Naive Bayes dalam klasifikasi sentimen, pemrosesan big data memerlukan teknologi yang mampu menangani data dalam jumlah besar. Sementara Naive Bayes dapat secara akurat mengklasifikasikan sentimen, Hadoop memungkinkan pemrosesan data yang efektif dalam skala besar. Bahkan ketika data tinjauan yang dipelajari mencapai jutaan entri, kombinasi Hadoop dan Naive Bayes meningkatkan akurasi analisis sentimen dan mempercepat waktu pemrosesan, menurut studi oleh [11].

Analisis sentimen yang didukung Hadoop pada platform e-commerce membantu bisnis lebih memahami preferensi dan keluhan pelanggan sekaligus meningkatkan efisiensi pemrosesan data besar. Temuan analisis dapat diterapkan untuk meningkatkan layanan pelanggan, kualitas produk, dan pembuatan strategi pemasaran. Dengan memanfaatkan Hadoop, strategi ini juga memungkinkan bisnis untuk menyesuaikan diri dengan perubahan di pasar.

DOI: 10.52362/jisicom.v9i1.1784

e-ISSN : 2597-3673 (Online) , p-ISSN : 2579-5201 (Printed)
Vol.9 No.1 (June 2025)
Journal of Information System, Informatics and Computing
Website/URL: http://journal.stmikjayakarta.ac.id/index.php/jisicom
Email: jisicom@stmikjayakarta.ac.id , jisicom2017@gmail.com

Oleh karena itu, tujuan dari penelitian ini adalah untuk menyelidiki bagaimana Hadoop dapat digunakan untuk menganalisis ulasan pelanggan pada platform e-commerce menggunakan algoritma Naive Bayes untuk analisis sentimen. Penelitian ini bertujuan untuk memberikan wawasan yang lebih baik kepada bisnis melalui pemrosesan Big Data sehingga mereka dapat meningkatkan pengalaman pelanggan dan bereaksi terhadap perkembangan pasar dengan lebih efektif.

II. TINJAUAN PUSTAKA

a. Analisis Sentimen

Tujuan dari analisis sentimen, subbidang pemrosesan bahasa alami, adalah untuk mengenali dan mengkategorikan perasaan atau pandangan yang diungkapkan dalam teks [12]. Karena penelitian sentimen dapat mengungkapkan bagaimana konsumen melihat dan terlibat dengan produk atau layanan tertentu, sangat penting untuk platform e-commerce. Analisis sentimen, misalnya, dapat digunakan untuk memprediksi apakah ulasan produk akan netral, menguntungkan, atau negatif. Temuan analisis dapat digunakan oleh bisnis untuk meningkatkan pengalaman pengguna, memodifikasi taktik pemasaran, dan menghasilkan produk berkualitas lebih tinggi.

Analisis sentimen untuk platform e-commerce telah banyak digunakan dalam penelitian sebelumnya. Sebuah studi oleh [13] yang menggunakan teknik klasifikasi sentimen untuk memeriksa ulasan pengguna di pasar dan e-commerce berfungsi sebagai salah satu contoh. Analisis sentimen dapat digunakan untuk mengontrol sentimen pelanggan dan menawarkan wawasan berharga tentang produk yang berkinerja baik atau sedang berjuang. Studi lain oleh [11] pada platform e-commerce Indonesia mengklaim bahwa analisis sentimen yang tepat dapat mempercepat kemampuan perusahaan untuk menyesuaikan diri dengan perubahan pasar.

b. Big Data dan Hadoop

Volume besar ulasan produk, berbagai format, dan seringkali data tidak terstruktur yang tidak dapat ditangani oleh perangkat lunak pemrosesan data tradisional atau sistem basis data adalah contoh Big Data di industri e-commerce. Ini menghadirkan kesulitan manajemen unik yang dapat diselesaikan dengan bantuan teknologi seperti Hadoop. Platform Hadoop sumber terbuka dibuat untuk menangani volume data yang sangat besar dengan cepat. Dengan bantuan MapReduce dan Hadoop distributed file systems (HDFS), ia dapat menyimpan dan memproses data dalam jumlah besar sekaligus [12].

Melalui penyebaran pemrosesan data di berbagai node dalam sebuah kluster, Hadoop memungkinkan pengolahan jumlah data yang sangat besar. Dalam penelitian yang dilakukan oleh [10], Hadoop digunakan untuk mengelola dan menganalisis ulasan pengguna di platform e-commerce. Penelitian tersebut menunjukkan bahwa Hadoop sangat efektif dalam mengelola volume data yang signifikan, meningkatkan efisiensi pengolahan data, dan menghasilkan wawasan yang lebih cepat dan akurat. Dengan menggunakan Hadoop, perusahaan dapat menemukan tren dan pola dalam ulasan pengguna dengan lebih mudah dan lebih cepat.

c. Analisis sentimen dengan Naïve Bayes

Salah satu metode pembelajaran mesin paling populer untuk analisis sentimen, terutama dalam klasifikasi teks, adalah Naive Bayes. Pendekatan ini bergantung pada teorema Bayes, yang menentukan probabilitas kelas berdasarkan fitur yang ditemukan dalam data. Asumsi bahwa setiap karakteristik dalam data bersifat independen digunakan oleh Naive Bayes, namun ini tidak selalu terjadi. Namun, Naive Bayes telah menunjukkan janji dalam sejumlah aplikasi [4].

Ulasan produk di situs e-*commerce* berhasil diklasifikasikan menggunakan Naive Bayes dalam sebuah penelitian oleh [5]. Studi ini menunjukkan bahwa bahkan dengan kumpulan data yang besar dan beragam, Naive Bayes masih dapat menghasilkan klasifikasi yang benar. Selanjutnya, [6] menemukan bahwa analisis sentimen ulasan produk di *e-commerce* Indonesia dapat dilakukan dengan menggunakan Naive Bayes. Temuan studi ini menunjukkan seberapa baik algoritma Naive Bayes mengklasifikasikan sentimen pengguna sebagai netral, negatif, atau negatif.

Dalam teks yang tidak terstruktur, seperti evaluasi produk, yang sering menggunakan bahasa santai dan beragam, Naive Bayes juga telah menunjukkan janji dalam mendeteksi sentimen. Penelitian oleh [13] menunjukkan bahwa analisis sentimen Naive Bayesian mengungguli teknik lain, seperti *Support Vector Machine* (SVM), dalam hal kecepatan dan akurasi.

DOI: 10.52362/jisicom.v9i1.1784

e-ISSN : 2597-3673 (Online) , p-ISSN : 2579-5201 (Printed)
Vol.9 No.1 (June 2025)
Journal of Information System, Informatics and Computing
Website/URL: http://journal.stmikjayakarta.ac.id/index.php/jisicom
Email: jisicom@stmikjayakarta.ac.id , jisicom2017@gmail.com

d. Integrasi Hadoop dan Naïve Bayes

Meskipun merupakan algoritma yang bagus untuk kategorisasi teks, Naive Bayes masih memiliki masalah dengan memproses data dalam jumlah besar, terutama ketika berhadapan dengan data yang sangat besar seperti yang terlihat dalam analisis ulasan *e-commerce*. Kombinasi Hadoop dan Naive Bayes, bagaimanapun, memberikan solusi: Hadoop dapat digunakan untuk pemrosesan terdistribusi dan penyimpanan data besar, sedangkan Naive Bayes dapat digunakan untuk klasifikasi dan analisis data besar [10].

Penelitian yang dilakukan oleh [12] menemukan bahwa penggabungan Hadoop dan Naive Bayes meningkatkan efisiensi pengolahan data besar. Dalam penelitian ini, Hadoop digunakan untuk melakukan pemrosesan data secara bersamaan, sedangkan Naive Bayes digunakan untuk melakukan klasifikasi sentimen terhadap ulasan produk. Hasil penelitian menunjukkan bahwa kombinasi keduanya dapat menghasilkan analisis sentimen yang lebih cepat dan lebih akurat daripada analisis Penggabungan ini juga memungkinkan perusahaan untuk membuat keputusan yang lebih berbasis data dan beradaptasi dengan perubahan pasar.

Perusahaan e-commerce dapat memperoleh pemahaman yang lebih baik tentang sentimen konsumen dan meresponsnya dengan lebih tepat waktu dengan integrasi pengolahan data besar yang efektif menggunakan Hadoop dan klasifikasi sentimen yang akurat menggunakan Naive Bayes. Kombinasi ini meningkatkan kinerja analisis sentimen dan memberikan perusahaan keuntungan kompetitif dalam pasar yang semakin kompetitif.

III. METODOLOGI PENELITIAN

Tujuan penelitian ini adalah untuk menggunakan teknologi Hadoop dan pendekatan Naive Bayes untuk mengelola data besar untuk menganalisis sentimen ulasan pelanggan di situs *e-commerce*. Metodologi penelitian ini dibagi menjadi beberapa fase utama, termasuk pengumpulan data, persiapan, pemrosesan data berbasis Hadoop, dan analisis sentimen menggunakan algoritma Naive Bayes. Berikut adalah rincian menyeluruh dari :

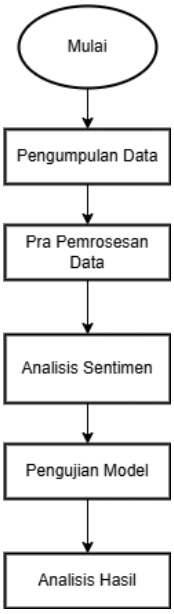
a. Pengumpulan Data

Data yang digunakan dalam penelitian ini berasal dari 391.500 ulasan Shopee yang diperoleh dengan menggunakan pengikisan Google Play Store. Data ulasan pengguna dikumpulkan dari platform e-commerce dengan pengikisan web menggunakan paket Python seperti BeautifulSoup atau Scrapy. Peringkat, konten ulasan, dan detail tambahan seperti tanggal ulasan atau kategori produk semuanya disertakan dalam data ini. Data ini disimpan dalam Hadoop Distributed File System (HDFS) untuk memfasilitasi pemrosesan data skala besar.

b. Pra-Pemrosesan Data

DOI: 10.52362/jisicom.v9i1.1784

e-ISSN : 2597-3673 (Online) , p-ISSN : 2579-5201 (Printed)
Vol.9 No.1 (June 2025)
Journal of Information System, Informatics and Computing
Website/URL: http://journal.stmikjayakarta.ac.id/index.php/jisicom
Email: jisicom@stmikjayakarta.ac.id , jisicom2017@gmail.com



Gambar 3. 1 Pra-Pemrosesan Data

c. Implementasi Hadoop

Teknologi *Apache Spark* digunakan untuk mengelola dan memproses volume data ulasan yang sangat besar. Dalam penelitian ini, *Hadoop Distributed File System* (HDFS) masih digunakan untuk penyimpanan data ulasan terdistribusi, tetapi PySpark digunakan untuk pemrosesan data. Tahapan yang dilakukan dengan Spark adalah sebagai berikut:

1. Penyimpanan Data di HDFS: Data yang telah diproses disimpan dalam HDFS, yang memungkinkan data didistribusikan ke berbagai node dalam kluster Hadoop.
2. Pemrosesan Data dengan PySpark: Data diproses menggunakan PySpark, yang memungkinkan data untuk didistribusikan ke berbagai node

d. Penerapan Naive Bayes untuk Analisis Sentimen

Setelah pemrosesan data PySpark, sentimen dari teks ulasan diklasifikasikan menggunakan metode Naive Bayes. Ada beberapa langkah yang terlibat dalam proses klasifikasi:

1. Distribusi Dataset: Data pelatihan dan data pengujian adalah dua komponen dari dataset tersebut. Data pengujian digunakan untuk mengevaluasi keakuratan model yang dilatih, sedangkan data pelatihan digunakan untuk melatih model Naive Bayes.
2. Ekstraksi Fitur: Model Naive Bayes menggunakan kata-kata dalam ulasan sebagai fitur. Setiap kata dianggap sebagai elemen yang berbeda dalam perhitungan perasaan potensial. Untuk menyiapkan vektor fitur untuk digunakan dalam model, operasi ekstraksi fitur dijalankan menggunakan Spark MLlib.
3. Pelatihan Model: Untuk menentukan kemungkinan sentimen berdasarkan kata-kata dalam ulasan, model Naive Bayes dilatih menggunakan kumpulan data pelatihan. Berdasarkan fitur (kata) dalam teks, algoritma Naive Bayes menggunakan PySpark MLlib untuk menentukan kemungkinan sentimen positif, negatif, atau netral.
4. Klasifikasi Sentimen: Setelah model dilatih, model Naive Bayes digunakan untuk mengklasifikasikan ulasan dalam dataset uji. Hasil klasifikasi akan menentukan sentimen dari ulasan, apakah positif, negatif, atau netral.

e. Evaluasi Model

Digunakan metrik evaluasi berikut untuk mengukur kemampuan model Naive Bayes untuk menganalisis sentimen:

1. Akurasi: Mengukur seberapa baik model mengklasifikasikan sentimen dengan benar.

DOI: 10.52362/jisicom.v9i1.1784

e-ISSN : 2597-3673 (Online) , p-ISSN : 2579-5201 (Printed)
Vol.9 No.1 (June 2025)
Journal of Information System, Informatics and Computing
Website/URL: http://journal.stmikjayakarta.ac.id/index.php/jisicom
Email: jisicom@stmikjayakarta.ac.id , jisicom2017@gmail.com

2. Presisi: Mengukur seberapa banyak ulasan yang diklasifikasikan sebagai positif, negatif, atau netral yang benar-benar sesuai dengan label yang seharusnya.
3. *Recall*: Mengukur seberapa banyak ulasan yang seharusnya diklasifikasikan dalam kategori tertentu benar-benar teridentifikasi.
4. F1-Score: Menunjukkan kinerja model secara keseluruhan melalui kombinasi recall dan presisi.

IV. PEMBAHASA DAN HASIL

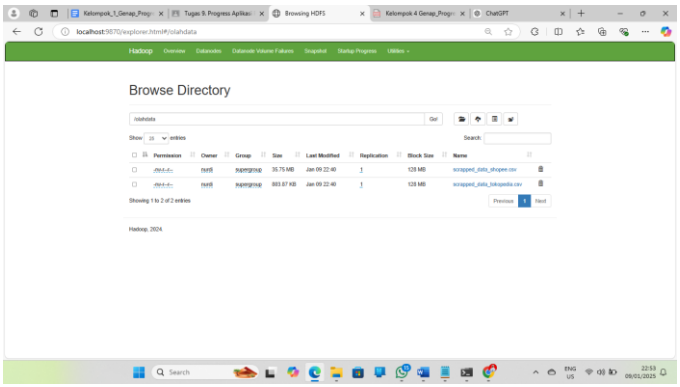
Langkah pertama dalam menggunakan Hadoop untuk analisis data adalah menginstal Apache Hadoop. Peneliti menggunakan Hadoop 3.2.1 dalam hal ini. Perintah "sudo wget -P ~ https://archive.apache.org/dist/hadoop/common/hadoop3.2.1/hadoop-3.2.1.tar.gz" dapat digunakan untuk menginstal Apache Hadoop di komputer Linux. Tunggu hingga unduhan selesai.

4.1. Menyiapkan Data

Data ulasan aplikasi Shopee yang diperoleh dari hasil scraping di Play Store disiapkan dengan bantuan Hadoop dan PySpark. Langkah pertama dalam menyiapkan data adalah mengunggah data ke Hadoop. Kode editor yang digunakan adalah Jupyter Notebook.

```
C:\Users\nurdi>hadoop fs -mkdir /olahdata

C:\Users\nurdi>hadoop fs -ls /
Found 3 items
drwxr-xr-x   - nurdi supergroup          0 2025-01-09 21:51 /bigdata
drwxr-xr-x   - nurdi supergroup          0 2025-01-09 22:38 /olahdata
drwxr-xr-x   - nurdi supergroup          0 2024-12-30 00:19 /user

C:\Users\nurdi>hadoop fs -put C:\Users\nurdi\Downloads\scrapped_data_shopee.csv /olahdata

C:\Users\nurdi>hadoop fs -put C:\Users\nurdi\Downloads\scrapped_data_tokopedia.csv /olahdata

C:\Users\nurdi>
```

Gambar 4. 1 Proses Pengunggahan Data ke Hadoop



Gambar 4. 2 Mengunggah Data ke Hadoop

Data dibersihkan dengan mengubah karakteristik dan tipe data setelah data diunggah ke Hadoop.

userName	score	at	content
Pengguna Google	5	2025-01-08 12:56:31	"Barang sudah nya...
Pengguna Google	5	2025-01-08 12:54:39	Shopee, bagaikan ...
Pengguna Google	5	2025-01-08 12:53:28	belanja lebih mudah
Pengguna Google	5	2025-01-08 12:52:49	Jos
Pengguna Google	5	2025-01-08 12:51:57	Aplikasi berbelan...

Gambar 4. 3 Tampilan data setelah disesuaikan atribut tipe datanya

DOI: 10.52362/jisicom.v9i1.1784

e-ISSN : 2597-3673 (Online) , p-ISSN : 2579-5201 (Printed)
Vol.9 No.1 (June 2025)
Journal of Information System, Informatics and Computing
Website/URL: http://journal.stmikjayakarta.ac.id/index.php/jisicom
Email: jisicom@stmikjayakarta.ac.id , jisicom2017@gmail.com

4.2. *Preprocessing Data*

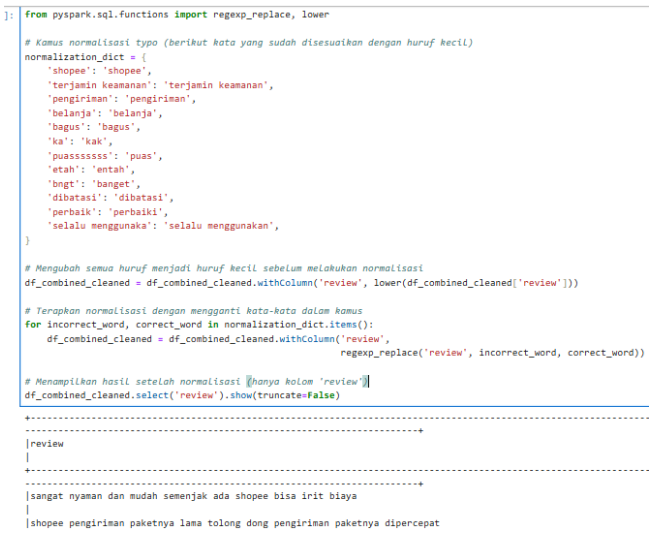
Langkah pertama dalam proses ini adalah menghapus item yang tidak perlu dari data. Misalnya, dengan menghilangkan tagar, sebutan, dan URL. Selain itu, karakter non-alfanumerik asing seperti simbol atau tanda baca tertentu juga ditinggalkan. Untuk membuat format data lebih rapi, spasi tambahan apa pun yang mungkin ada dalam data seperti di awal dan akhir teks juga dihilangkan.

```
[13]: from pyspark.sql.functions import col, regexp_replace

# Fungsi untuk membersihkan teks
def clean_text(df, column_name):
    # Hapus URL
    df = df.withColumn(column_name, regexp_replace(col(column_name), r'http\S+|www\S+', ''))

    # Hapus mention (misalnya @username)
    df = df.withColumn(column_name, regexp_replace(col(column_name), r'@\S+', ''))

    # Hapus hashtag (misalnya #hashtag)
    df = df.withColumn(column_name, regexp_replace(col(column_name), r'#\S+', ''))

    # Hapus karakter non-alfanumerik kecuali spasi
    df = df.withColumn(column_name, regexp_replace(col(column_name), r'[^a-zA-Z0-9\s]', ''))

    # Hapus spasi berlebihan
    df = df.withColumn(column_name, regexp_replace(col(column_name), r'\s+', ' '))

    # Hapus spasi di awal dan akhir teks
    df = df.withColumn(column_name, regexp_replace(col(column_name), r'^\s+|\s+?$', ''))

    return df

# Terapkan fungsi untuk membersihkan teks pada kolom 'review'
df_combined_cleaned = clean_text(df_combined, 'review')

# Tampilkan sampel data setelah preprocessing
df_combined_cleaned.show(5)
```

```
+------+-------------------+--------------------+
|rating|                 at|              review|
+------+-------------------+--------------------+
|     5|2025-01-08 08:21:07|Sangat nyaman dan...|
|     1|2025-01-08 05:10:05|Shopee pengiriman...|
|     5|2025-01-08 03:16:34|                good|
|     1|2025-01-07 15:39:26|Pesanan selalu di...|
|     5|2025-01-07 13:17:18|Pengiriman cepat ...|
+------+-------------------+--------------------+
```

Gambar 4. 4 Tahap Preprocessing data

Setelah tahap *preprocessing* data, proses selanjutnya adalah mengganti kata-kata dengan ejaan atau makna yang tidak sesuai menggunakan kamus normalisasi (*normalization dictionary*). Kamus ini berisi pasangan kata yang sering salah eja atau tidak baku (misalnya kata slang atau singkatan), yang kemudian diganti dengan kata yang benar dan sesuai konteks. Langkah ini dilakukan untuk meningkatkan kualitas teks ulasan sehingga lebih terstruktur dan mudah dianalisis. Dalam implementasi ini, setiap kata dalam kolom review yang sesuai dengan entri di kamus akan digantikan dengan kata yang lebih formal atau standar.



```
]: from pyspark.sql.functions import regexp_replace, lower

# Kamus normalisasi typo (berikut kata yang sudah disesuaikan dengan huruf kecil)
normalization_dict = {
    'shopee': 'shopee',
    'terjamin keamanan': 'terjamin keamanan',
    'pengiriman': 'pengiriman',
    'belanja': 'belanja',
    'bagus': 'bagus',
    'ka': 'kak',
    'puasssssss': 'puas',
    'etah': 'entah',
    'bngt': 'banget',
    'dibatasi': 'dibatasi',
    'perbaik': 'perbaiki',
    'selalu mengguna': 'selalu menggunakan',
}

# Mengubah semua huruf menjadi huruf kecil sebelum melakukan normalisasi
df_combined_cleaned = df_combined_cleaned.withColumn('review', lower(df_combined_cleaned['review']))

# Terapkan normalisasi dengan mengganti kata-kata dalam kamus
for incorrect_word, correct_word in normalization_dict.items():
    df_combined_cleaned = df_combined_cleaned.withColumn('review',
                                                         regexp_replace('review', incorrect_word, correct_word))

# Menampilkan hasil setelah normalisasi (hanya kolom 'review')
df_combined_cleaned.select('review').show(truncate=False)
```

```
+--------------------------------------------------------------------------------------------------------------------------------------------------------------------+
|review                                                                                                                                                              |
+--------------------------------------------------------------------------------------------------------------------------------------------------------------------+
|sangat nyaman dan mudah semenjak ada shopee bisa irit biaya                                                                                                         |
|shopee pengiriman paketnya lama tolong dong pengiriman paketnya dipercepat                                                                                          |
```

Gambar 4. 5 Normalisasi Data

DOI: 10.52362/jisicom.v9i1.1784

e-ISSN : 2597-3673 (Online) , p-ISSN : 2579-5201 (Printed)
Vol.9 No.1 (June 2025)
Journal of Information System, Informatics and Computing
Website/URL: http://journal.stmikjayakarta.ac.id/index.php/jisicom
Email: jisicom@stmikjayakarta.ac.id , jisicom2017@gmail.com

Setelah proses normalisasi selesai, langkah selanjutnya adalah tokenisasi. Tokenisasi adalah proses memecah teks dalam kolom review menjadi unit-unit kata atau token. Proses ini bertujuan agar setiap kata dalam ulasan dapat dianalisis secara individu, memungkinkan analisis lebih mendalam pada setiap komponen teks.

```
from pyspark.ml.feature import Tokenizer, StopWordsRemover

# Tokenisasi ulasan teks
tokenizer = Tokenizer(inputCol="content", outputCol="words")
df_tokens = tokenizer.transform(df_labelled)

# Menghapus stopwords (kata-kata umum yang tidak mengandung informasi penting)
remover = StopWordsRemover(inputCol="words", outputCol="filtered_words")
df_filtered = remover.transform(df_tokens)

# Menampilkan hasil tokenisasi dan penghapusan stopwords
df_filtered.select("content", "filtered_words").show(5)
```

```
+--------------------+--------------------+
|             content|      filtered_words|
+--------------------+--------------------+
|"Barang sudah nya...|["barang, sudah, ...|
|Shopee, bagaikan ...|[shopee,, bagaika...|
| belanja lebih mudah|[belanja, lebih, ...|
|                 Jos|               [jos]|
|Aplikasi berbelan...|[aplikasi, berbel...|
+--------------------+--------------------+
only showing top 5 rows
```

Gambar 4. 6 Tokenisasi data

Menghilangkan stopwords adalah langkah penting dalam pemrosesan teks yang bertujuan untuk meningkatkan kualitas analisis data. Stopwords adalah kata-kata umum seperti "dan," "di," "yang," dan sejenisnya, yang sering muncul dalam teks tetapi memiliki kontribusi makna yang kecil dalam analisis. Dengan menghapus stopwords, fokus analisis dapat diarahkan pada kata-kata yang lebih relevan dan bermakna. Sebagai contoh, kalimat "saya suka makan di restoran" setelah dihapus stopwords menjadi "suka makan restoran." Langkah ini dilakukan menggunakan fungsi StopWordsRemover, di mana kolom yang berisi token akan diproses untuk menghasilkan kolom baru dengan token yang telah disaring. Proses ini membantu mempermudah analisis lebih lanjut seperti penghitungan frekuensi kata atau penerapan model pembelajaran mesin.

```
# Langkah 2: Menghilangkan stopwords - Menghapus kata yang tidak penting dari tokens
remover = StopWordsRemover(inputCol="tokens", outputCol="filtered_tokens")
df_combined_cleaned = remover.transform(df_combined_cleaned)

# Menampilkan hasil setelah menghilangkan stopwords
df_combined_cleaned.select('filtered_tokens').show(truncate=False)
```

```
+-----------------------------------------------------------------------------------------------
------------------------------------------------------------------------------------------+
|filtered_tokens
|
+-----------------------------------------------------------------------------------------------
------------------------------------------------------------------------------------------+
|[sangat, nyaman, dan, mudah, semenjak, ada, shopee, bisa, irit, biaya]
|
|[shopee, pengiriman, paketnya, lama, tolong, dong, pengiriman, paketnya, dipercepat]
|
|[good]
```

Gambar 4. 7 Menghapus stopwords

Setelah proses pembersihan data selesai, langkah selanjutnya adalah memilih kolom yang relevan untuk analisis sentimen. Dalam kasus ini, kolom yang dipilih mencakup rating, tanggal ulasan (at), dan filtered_tokens_string. Kolom rating digunakan sebagai label sentimen yang akan dianalisis, sementara filtered_tokens_string berisi token teks ulasan yang telah difilter dan siap untuk diproses lebih lanjut. Pemilihan kolom ini bertujuan untuk menyederhanakan dataset sehingga hanya memuat informasi yang benar-benar diperlukan untuk analisis. Dengan langkah ini, data menjadi lebih terfokus dan siap untuk digunakan pada tahap

DOI: 10.52362/jisicom.v9i1.1784

e-ISSN : 2597-3673 (Online) , p-ISSN : 2579-5201 (Printed)
Vol.9 No.1 (June 2025)
Journal of Information System, Informatics and Computing
Website/URL: http://journal.stmikjayakarta.ac.id/index.php/jisicom
Email: jisicom@stmikjayakarta.ac.id , jisicom2017@gmail.com

analisis sentimen atau pelatihan model pembelajaran mesin. Langkah terakhir dalam proses ini adalah menampilkan dataset yang sudah disederhanakan untuk memastikan bahwa kolom yang dipilih sudah sesuai dengan kebutuhan analisis.

```
# Memilih kolom yang dibutuhkan untuk analisis sentimen
df_sentimen = df_combined_cleaned.select('rating', 'at', 'filtered_tokens_string')

# Menampilkan hasil untuk memastikan kolom yang dipilih
df_sentimen.show(truncate=False)
```

```
+------+-------------------+--------------------------------------------------------------------------------
|rating|at                 |filtered_tokens_string
|
+------+-------------------+--------------------------------------------------------------------------------
|5     |2025-01-08 08:21:07|sangat nyaman dan mudah semenjak ada shopee bisa irit biaya
|
|1     |2025-01-08 05:10:05|shopee pengiriman paketnya lama tolong dong pengiriman paketnya dipercepat
|
|5     |2025-01-08 03:16:34|good
|
|1     |2025-01-07 15:39:26|pesanan selalu dibatalkakn otomatis padahal gak ngapa2in koq makin aneh yah shopee skrg
|
|5     |2025-01-07 13:17:18|pengiriman cepat proses pengembalian barang atau dana yang ngga sesuap pesanan atau foto sangat mudah
|
|5     |2025-01-06 18:00:01|saya sebagai user pembeli makin kesini shopee lebih banyak varian barangnya serta bagus baik dalam support dan proses refaun
d lebih cepat dibandingkakn si hijau daun lokakl lebih menyenangkakn pembeli dibandingkakn pelapak|
|5     |2025-01-06 12:55:25|terjamin keamanannya belanja disini
|
|5     |2025-01-06 12:08:10|bagus
|
|4     |2025-01-06 09:30:36|
|
|5     |2025-01-06 08:47:10|belanja jadi mudah
```

Gambar 4. 8 Memilih kolom yang relevan

3.3. Pelabelan Data

Langkah berikutnya adalah menambahkan kolom baru bernama label untuk mengkategorikan sentimen berdasarkan nilai rating. Proses ini dilakukan menggunakan fungsi kondisional untuk menentukan label sentimen. Aturannya adalah sebagai berikut:

a. Jika nilai score adalah **5**, maka label sentimen diberi nilai **1** (positif).
b. Jika nilai score tidak sama dengan **5** (misalnya, 1, 2, 3, atau 4), maka label sentimen diberi nilai **0** (negatif).

Dengan menambahkan kolom label, setiap ulasan akan diberi kategori sentimen yang dapat digunakan untuk analisis lebih lanjut, seperti eksplorasi data atau pelatihan model klasifikasi. Setelah kolom ini ditambahkan, dataset ditampilkan kembali untuk memastikan bahwa penambahan kolom label sudah dilakukan dengan benar sesuai dengan logika yang diimplementasikan.

```
[6]: from pyspark.sql.functions import when

# Menambahkan kolom baru 'sentiment' berdasarkan kolom 'score'
df_labelled = df_clean.withColumn(
    "sentiment",
    when(df_clean['score'] == 5, 1).otherwise(0)  # 1 untuk positif, 0 untuk negatif
)

df_labelled.select("userName", "score", "sentiment", "content").show(5)
```

```
+---------------+-----+---------+--------------------+
|       userName|score|sentiment|             content|
+---------------+-----+---------+--------------------+
|Pengguna Google|    5|        1|"Barang sudah nya...|
|Pengguna Google|    5|        1|Shopee, bagaikan ...|
|Pengguna Google|    5|        1| belanja lebih mudah|
|Pengguna Google|    5|        1|                 Jos|
|Pengguna Google|    5|        1|Aplikasi berbelan...|
+---------------+-----+---------+--------------------+
only showing top 5 rows
```

Gambar 4. 9 Pelabelan data

3.4. Pemodelan Sentimen

Dalam penelitian ini, model kategorisasi sentimen dibangun dengan menggunakan teknik Naive Bayes. Dua subset data dibuat: 20% untuk data uji dan 80% untuk data pelatihan. Untuk menjamin distribusi data yang

DOI: 10.52362/jisicom.v9i1.1784

e-ISSN : 2597-3673 (Online) , p-ISSN : 2579-5201 (Printed)
Vol.9 No.1 (June 2025)
Journal of Information System, Informatics and Computing
Website/URL: http://journal.stmikjayakarta.ac.id/index.php/jisicom
Email: jisicom@stmikjayakarta.ac.id , jisicom2017@gmail.com

seimbang, fungsi randomSplit pustaka PySpark digunakan selama proses berbagi data. Menggunakan parameter featuresCol sebagai representasi fitur, labelCol sebagai label sentimen, dan jenis model multinomial yang tepat untuk distribusi data teks, model Naive Bayes dilatih.

```
# Membagi data menjadi 80% data latih dan 20% data uji
train_data, test_data = df_vectorized.randomSplit([0.8, 0.2], seed=42)
```

```
from pyspark.ml.classification import NaiveBayes

# Membangun model Naive Bayes
nb = NaiveBayes(featuresCol="features", labelCol="sentiment", modelType="multinomial")

# Melatih model Naive Bayes dengan data latih
nb_model = nb.fit(train_data)

# Menyimpan model di E:/Big Data
nb_model.save("hdfs://localhost:9000/olahdata/naive_bayes_model")
```

Gambar 4. 10 Pemodelan sentimen

Gambar 10 menunjukkan alur proses pembentukan model, mulai dari pembagian data hingga penyimpanan model ke dalam sistem file terdistribusi Hadoop (HDFS). Penyimpanan model dilakukan di direktori hdfs://localhost:9000/olahdata/naive_bayes_model, yang mempermudah pengelolaan dan implementasi ulang model pada skala data besar. Langkah ini bertujuan untuk mengoptimalkan efisiensi dalam pengolahan data besar menggunakan Hadoop dan PySpark. Pada tahap evaluasi, model ini akan dianalisis lebih lanjut untuk mengukur performa klasifikasinya. Evaluasi tersebut mencakup metrik seperti akurasi, presisi, dan recall, yang akan dijelaskan pada bagian berikutnya.

Data terlatih yang telah diproses digunakan untuk melatih model Naive Bayes setelah data dibagi menjadi 80% data pelatihan dan 20% data uji. Dengan menggunakan data pengujian yang tidak digunakan untuk pelatihan, kinerja model kemudian dievaluasi. Tujuan dari pengujian ini adalah untuk mengevaluasi kapasitas model untuk menggeneralisasi ke data baru. Metrik akurasi digunakan untuk mengukur kebenaran model sebagai bagian dari proses evaluasi. Dengan membandingkan temuan prediksi model dengan label asli data pengujian (sentimen), akurasi ditentukan. Tingkat akurasi {akurasi} dihasilkan oleh model Naive Bayes berdasarkan evaluasi yang dilakukan

```
from pyspark.sql.functions import col

# Menghitung jumlah prediksi benar
correct_predictions = predictions.filter(col("sentiment") == col("final_prediction")).count()

# Total data
total_data = predictions.count()

# Akurasi
accuracy = correct_predictions / total_data
print(f"Akurasi model: {accuracy * 100:.2f}%")
```

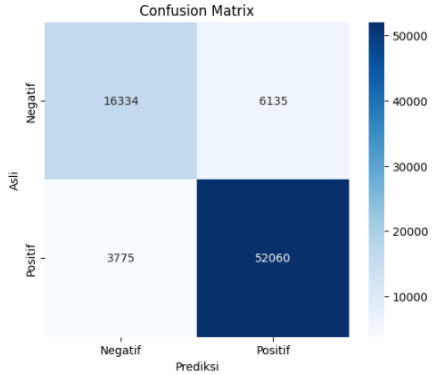
```
Akurasi model: 88.62%
```

Gambar 4. 11 Hasil akurasi

Gambar 11 memberikan ilustrasi proses evaluasi model, mulai dari pembagian data, pelatihan model, hingga pengujian dan perhitungan metrik akurasi. Temuan evaluasi menunjukkan bahwa, mengingat data yang ada, model berkinerja baik dalam klasifikasi sentimen. Metrik tambahan seperti presisi, penarikan, dan skor F1 harus dihitung, khususnya untuk memahami performa model pada setiap kategori sentimen, untuk mendapatkan gambaran yang lebih menyeluruh tentang performa model.

DOI: 10.52362/jisicom.v9i1.1784

e-ISSN : 2597-3673 (Online) , p-ISSN : 2579-5201 (Printed)
Vol.9 No.1 (June 2025)
Journal of Information System, Informatics and Computing
Website/URL: http://journal.stmikjayakarta.ac.id/index.php/jisicom
Email: jisicom@stmikjayakarta.ac.id , jisicom2017@gmail.com



Gambar 4. 12 *Confusion matrix*

Gambar 12 menunjukkan matriks kebingungan (*confusion matrix*) yang digunakan untuk mengevaluasi hasil klasifikasi model Naive Bayes pada data uji. Matriks ini menggambarkan performa model dalam mengklasifikasikan sentimen, dengan rincian sebagai berikut: sebanyak 16.334 data sentimen negatif berhasil diklasifikasikan dengan benar sebagai negatif (True Negatives), sementara 6.135 data sentimen negatif salah diklasifikasikan sebagai positif (False Positives). Di sisi lain, model berhasil mengklasifikasikan 52.060 data sentimen positif dengan benar sebagai positif (True Positives), namun terdapat 3.775 data sentimen positif yang salah diklasifikasikan sebagai negatif (False Negatives). Hasil ini menunjukkan bahwa model memiliki kinerja yang lebih baik dalam mengklasifikasikan sentimen positif dibandingkan dengan sentimen negatif, yang terlihat dari jumlah True Positives yang jauh lebih tinggi dibandingkan True Negatives.

Tabel 4. 1 *Classification report*

	precission	*Recall*	*F1-socre*	*Support*
Negatif	0.81	0.73	0.77	22469
Positif	0.89	0.93	0.91	55835
accuracy			0.87	78304
Macro avg	0.85	0.83	0.84	78304
Weighted avg	0.87	0.87	0.87	78304

Pada Tabel 4.1 ditampilkan laporan evaluasi kinerja model berdasarkan metrik utama seperti precision, recall, dan F1-score. Precision untuk sentimen negatif adalah 0,81, yang berarti 81% prediksi negatif benar, sementara precision untuk sentimen positif mencapai 0,89, menunjukkan 89% prediksi positif sesuai. Recall untuk sentimen negatif sebesar 0,73 mengindikasikan 73% data negatif berhasil diklasifikasikan dengan benar sebagai negatif, sedangkan recall untuk sentimen positif mencapai 0,93, menunjukkan kemampuan model yang sangat baik dalam mengenali data positif. F1-score, yang menggabungkan precision dan recall untuk memberikan metrik kinerja yang seimbang, tercatat pada nilai 0,77 untuk sentimen negatif dan 0,91 untuk sentimen positif. Selain itu, akurasi keseluruhan model mencapai 87%, menunjukkan bahwa model berhasil mengklasifikasikan 87% data uji dengan benar. Rata-rata makro menunjukkan nilai precision 0,85, recall 0,83, dan F1-score 0,84, yang menggambarkan performa rata-rata model tanpa memperhatikan distribusi data. Sementara itu, rata-rata tertimbang untuk precision, recall, dan F1-score masing-masing bernilai 0,87, yang mencerminkan performa model dengan mempertimbangkan distribusi data pada setiap kelas. Hasil ini menunjukkan bahwa model memiliki kinerja yang baik dalam mengklasifikasikan sentimen, terutama pada data sentimen positif.

3.5. Visualisasi

Metode ini membantu mengidentifikasi kata-kata yang lebih relevan untuk analisis sentimen. Kata-kata yang memiliki bobot lebih tinggi dalam perhitungan TF-IDF sering kali mencerminkan opini atau sentimen

DOI: 10.52362/jisicom.v9i1.1784

e-ISSN : 2597-3673 (Online) , p-ISSN : 2579-5201 (Printed)
Vol.9 No.1 (June 2025)
Journal of Information System, Informatics and Computing
Website/URL: http://journal.stmikjayakarta.ac.id/index.php/jisicom
Email: jisicom@stmikjayakarta.ac.id , jisicom2017@gmail.com

pengguna yang lebih signifikan, sehingga memberikan informasi tambahan yang berguna pada proses analisis. Setelah data dikonversi ke dalam bentuk numerik, Langkah selanjutnya adalah membuat visualisasi berupa *wordcloud*. Visualisasi ini dilakukan untuk memberikan gambaran mengenai kata-kata yang lebih sering muncul dalam data ulasan.

Gambar 4. 13 *Wordcloud* Ulasan Positif

Pada Gambar 13 menampilkan *Word Cloud* yang menggambarkan distribusi kata-kata dominan pada data sentimen positif. Kata-kata seperti "sangat membantu," "di shopee," "belanja di," dan "gratis ongkir" muncul dengan ukuran font yang lebih besar, menunjukkan frekuensi kemunculannya yang tinggi dalam ulasan positif. Hal ini mengindikasikan bahwa aspek kemudahan belanja, ketersediaan promo seperti gratis ongkir, serta kualitas layanan menjadi faktor utama yang mendorong sentimen positif dari pengguna. Selain itu, kata-kata seperti "puas," "mudah," dan "sangat baik" mencerminkan apresiasi pengguna terhadap pengalaman belanja online yang efisien dan memuaskan melalui aplikasi Shopee. *Word Cloud* ini memberikan wawasan visual yang berguna dalam mengidentifikasi faktor-faktor penting yang berkontribusi pada pengalaman positif pengguna. Dengan informasi ini, pihak pengelola dapat memprioritaskan aspek-aspek yang dihargai pelanggan, seperti mempertahankan program promo dan meningkatkan fitur aplikasi yang mendukung kemudahan penggunaan, untuk memastikan tingkat kepuasan yang tinggi tetap terjaga.

Gambar 4. 14 *Wordcloud* Ulasan Negatif

DOI: 10.52362/jisicom.v9i1.1784

e-ISSN : 2597-3673 (Online) , p-ISSN : 2579-5201 (Printed)
Vol.9 No.1 (June 2025)
Journal of Information System, Informatics and Computing
Website/URL: http://journal.stmikjayakarta.ac.id/index.php/jisicom
Email: jisicom@stmikjayakarta.ac.id , jisicom2017@gmail.com

Pada Gambar 14 menampilkan *wordcloud* sentimen negatif, keluhan pengguna terhadap layanan Shopee menunjukkan berbagai persoalan yang mendesak untuk diatasi, terutama terkait performa aplikasi, sistem pembayaran, dan layanan pelanggan. Kata-kata seperti "tidak bisa" dan "gak bisa" muncul dengan sangat menonjol, mencerminkan adanya masalah signifikan yang membuat pengguna merasa terhambat dalam menggunakan layanan. Masalah-masalah ini kemungkinan besar berkaitan dengan fitur yang tidak berfungsi sebagaimana mestinya, seperti proses pembayaran, pengiriman barang, atau pengelolaan saldo pada akun pengguna. Selain itu, kata "lagi" yang sering muncul mengindikasikan adanya keluhan berulang dari pengguna, yang mungkin merasa frustrasi karena masalah yang sama terus terjadi tanpa penyelesaian yang memuaskan. Kata "juga" dan "aja" menggarisbawahi kesan bahwa pengguna merasa terbatas dalam pilihan atau solusi yang tersedia, sehingga menciptakan pengalaman negatif secara keseluruhan.

Penggunaan kata "tolong" dalam ulasan ini menunjukkan adanya tingginya permintaan akan bantuan atau dukungan dari pihak Shopee. Ini bisa menjadi cerminan dari sistem layanan pelanggan yang dirasa kurang responsif atau tidak mampu memberikan solusi yang memadai terhadap permasalahan pengguna. Kata "saya" juga sering muncul, menandakan bahwa pengguna seringkali berbicara dari perspektif pribadi dan menyuarakan kekecewaan mereka secara langsung. Kata-kata seperti "itu" dan "karena" memberikan konteks tambahan pada keluhan, menunjukkan bahwa pengguna mencoba memberikan penjelasan terkait permasalahan yang mereka alami. Sementara itu, kehadiran kata "masih" dapat menjadi indikasi adanya persoalan yang belum terselesaikan atau dirasakan terus berlanjut meskipun telah dilaporkan.

Semua hal dipertimbangkan, evaluasi yang tidak menguntungkan ini menunjukkan berapa banyak area penawaran Shopee yang memerlukan peningkatan serius. Memulihkan kepercayaan dan kebahagiaan pengguna membutuhkan sejumlah tindakan penting, termasuk meningkatkan kinerja aplikasi untuk mengurangi bug atau masalah teknis, meningkatkan transparansi dalam sistem pembayaran dan manajemen transaksi, serta memberikan dukungan pelanggan yang lebih cepat dan efektif. Mengingat persaingan yang semakin ketat di industri e-commerce, shopee harus melihat keluhan ini sebagai umpan balik yang berguna untuk meningkatkan kualitas layanan dan menjaga citra platform.

V. KESIMPULAN

Informasi yang digunakan dalam penelitian ini diperoleh dengan memanen 391.500 ulasan dari aplikasi Shopee dari Google Play Store. Temuan penelitian menunjukkan bahwa, dengan tingkat akurasi 87%, penggunaan algoritma Naive Bayes oleh Hadoop telah berhasil menilai nada ulasan pengguna di situs e-commerce. Dengan skor F1 0,91 untuk sentimen positif dan 0,77 untuk sentimen negatif, model klasifikasi berkinerja berbeda untuk setiap sentimen. Studi ini menggunakan analisis word cloud untuk menunjukkan bahwa pendorong utama sentimen positif di antara konsumen Shopee adalah hal-hal seperti pengiriman gratis, keramahan pengguna, dan penawaran khusus. Sementara itu, masalah teknis dengan aplikasi, sistem pembayaran, dan responsivitas dukungan pelanggan adalah kekhawatiran utama pengguna yang ditemukan melalui penelitian sentimen negatif.

Temuan studi ini menunjukkan nilai penerapan teknologi big data untuk analisis sentimen di industri e-commerce dan menawarkan dasar untuk penelitian masa depan semacam ini. Hasil ini dapat berfungsi sebagai dasar untuk pengembangan fitur dan layanan platform *e-commerce*, dengan penekanan pada area di mana pengguna sering mengungkapkan ketidakpuasan. Anda menjelaskan hasil yang diukur atau diuji dalam diskusi artikel penelitian, serta apa yang dicapai dan bagaimana hal itu memajukan sains dan mempersiapkan jalan untuk studi di masa depan.

REFERENASI

[1] N. Fatimah Az-Zahrah, R. Putra Dwitama, A. J. Suryaputra, F. Rahma, and J. Informatika, “Dampak E-commerce Terhadap Bidang Ekonomi, Bisnis, dan Pembelajaran: Tinjauan Literatur,” *Jurnal Teknologi Informasi*, vol. 4, no. 2, 2023, doi: 10.46576/djtechno.

[2] Y. L. R. Rehatalanit, “Peran E-commerce dalam Pengembangan Bisnis,” 2020.

[3] N. K. Aula and S. Suharto, “Pengaruh e-commerce terhadap Produk Domestik Bruto Indonesia,” *Jurnal Kebijakan Ekonomi dan Keuangan*, vol. 1, no. 1, pp. 39–48, Jun. 2021, doi: 10.20885/jkek.vol1.iss1.art4.

DOI: 10.52362/jisicom.v9i1.1784

e-ISSN : 2597-3673 (Online) , p-ISSN : 2579-5201 (Printed)
Vol.9 No.1 (June 2025)
Journal of Information System, Informatics and Computing
Website/URL: http://journal.stmikjayakarta.ac.id/index.php/jisicom
Email: jisicom@stmikjayakarta.ac.id , jisicom2017@gmail.com

[4] Rahel Lina Simanjuntak, Theresia Romauli Siagian, Vina Anggriani, and Arnita Arnita, "Analisis Sentimen Ulasan Pada Aplikasi E-Commerce Shopee Dengan Menggunakan Algoritma Naïve Bayes," *Jurnal Teknik Mesin, Elektro dan Ilmu Komputer*, vol. 3, no. 3, pp. 23–39, Nov. 2023, doi: 10.55606/teknik.v3i3.2411.

[5] A. Muzaki *et al.*, "Analisis Sentimen pada Ulasan Produk di E-Commerce dengan Metode Naive Bayes," *Jurnal Riset dan Aplikasi Mahasiswa Informatika (JRAMI)*, vol. 05, no. 04, 2024.

[6] A. H. Hasugian, M. Fakhriza, and D. Zukhoiriyah, "Analisis Sentimen pada Review Pengguna E-Commerce Menggunakan Algoritma Naive Bayes," *Januari*, no. 1, pp. 98–107, 2023, [Online]. Available: https://ojs.trigunadharma.ac.id/index.php/jsk/index

[7] N. Arif Maulana and Z. Fatah, "Penerapan Metode Naive Bayes untuk Analisis Sentimen Ulasan Produk di Platfrom E-Commerce," *Gudang Jurnal Multidisiplin Ilmu*, vol. 2, pp. 433–439, 2024, doi: 10.59435/gjmi.v2i11.1103.

[8] S. Yang, "PERNYATAAN BEBAS PLAGIARISME."

[9] A. H. Hasugian, M. Fakhriza, and D. Zukhoiriyah, "Volume 6 ; Nomor 1," *Januari*, 2023, [Online]. Available: https://ojs.trigunadharma.ac.id/index.php/jsk/index

[10] N. M. Y. D. A. Ni Made Yulia Dewati Ayu and Jakaria, "Penngaruh E-commerce Terhadap Pertumbuhan Ekonomi Indonesia," *Jurnal Ekonomi Trisakti*, vol. 3, no. 2, pp. 2891–2900, Aug. 2023, doi: 10.25105/jet.v3i2.17499.

[11] A. Firmansyah, "Kajian Kendala Implementasi E-commerce di Indonesia," 2020.

[12] P. D. Atika, P. D. Atika, and S. Suhadi, "Implementasi Algoritma Naïve Bayes Classifier untuk Analisis Sentimen Customer pada Toko Online," *Faktor Exacta*, vol. 12, no. 4, p. 303, Feb. 2020, doi: 10.30998/faktorexacta.v12i4.5224.

[13] M. Pradana, "Klasifikasi Bisnis E-commerce di Indonesia," *163 MODUS*, vol. 27, no. 2, p. 2, 2020.

DOI: 10.52362/jisicom.v9i1.1784