
ANALISIS SENTIMEN ULASAN SHOPEE PADA GOOGLE *PLAY STORE* MENGGUNAKAN ALGORITMA NAIVE BAYES

Riki Roska Rismansyah¹⁾, Aso Sudiarjo²⁾, Teuku Mufizar³⁾

^{1,2,3}Prodi Teknik Informatika, Fakultas Teknik, Universitas Perjuangan Tasikmalaya

Correspondence author: R.R.Rismansyah, 2203010606@unper.ac.id, Tasikmalaya, Indonesia

Abstract

This study aims to explore user opinions on the Shopee e-commerce application through sentiment analysis of reviews on the Google Play Store. Using the Python Programming Language with web scraping techniques, 2000 review data were collected and processed using the Naïve Bayes algorithm. The data preprocessing includes text cleaning, tokenization, and removal of stop words. The classification results show an accuracy of 83%, with a dominance of negative sentiment. From the results of the graphical visualization of consumer data reviews, negative sentiment data reviews get 65.9%, and positive sentiment data reviews get 34.1%. This study contributes to developing a sentiment-based recommendation system and provides input for Shopee to improve service quality.

Keywords: sentiment analysis, naïve bayes, web scraping, shopee

Abstrak

Penelitian ini bertujuan untuk menggali opini pengguna terhadap aplikasi *e-commerce* Shopee melalui analisis sentimen terhadap ulasan yang terdapat di Google Play Store. Menggunakan Bahasa Pemrograman Python dengan teknik *web scraping*, sebanyak 2000 data ulasan berhasil dihimpun dan diolah menggunakan algoritma Naïve Bayes. Proses *preprocessing* data meliputi pembersihan teks, tokenisasi, dan penghilangan *stop words*. Hasil klasifikasi menunjukkan akurasi sebesar 83%, dengan dominasi sentimen negatif. Dari hasil visualisasi ulasan data konsumen secara grafis, ulasan data sentimen negatif mendapatkan presentase sebanyak 65.9% dan untuk ulasan data sentiment positif mendapatkan presentase sebanyak 34.1%. Penelitian ini memberikan kontribusi dalam pengembangan sistem rekomendasi berbasis sentimen serta memberikan masukan bagi pihak Shopee untuk meningkatkan kualitas layanan.

Kata Kunci: analisis sentimen, ulasan, *Play Store*, shopee, *web scraping*, *naïve bayes*

A. PENDAHULUAN

Seiring pesatnya perkembangan teknologi digital, aplikasi *e-commerce* seperti Shopee telah menjadi bagian tak terpisahkan dari kehidupan sehari-hari

pengguna *smartphone*. Kemudahan berbelanja secara online telah membuat aplikasi ini sangat populer di Asia Tenggara (Oktia et al., 2024). Dalam persaingan aplikasi *e-commerce* yang semakin ketat, Shopee telah berhasil memikat jutaan

pengguna di Indonesia. Namun, untuk mempertahankan posisi terdepan, Shopee perlu secara proaktif mendengarkan suara pelanggan (Asih, 2024). Melalui analisis sentimen terhadap ulasan pengguna di Play Store, Shopee dapat mengidentifikasi secara mendalam apa yang disukai dan tidak disukai pengguna terhadap aplikasi mereka. Dengan demikian, Shopee dapat mengambil langkah-langkah konkret untuk meningkatkan kualitas layanan, seperti memperbaiki fitur yang sering dikeluhkan, meningkatkan kecepatan loading aplikasi, atau memberikan penawaran yang lebih menarik (Agustina et al., 2022). Analisis sentimen ini juga dapat membantu Shopee dalam mengembangkan fitur-fitur baru yang sesuai dengan kebutuhan dan ekspektasi pengguna (Nurian et al., 2024).

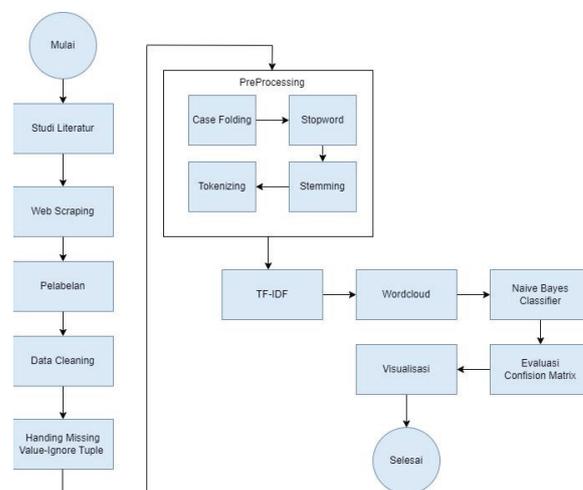
Dalam era digital, ulasan pengguna telah menjadi faktor penentu dalam keputusan pembelian konsumen. Jumlah ulasan yang melimpah di berbagai platform, seperti Play Store, membuat perusahaan kesulitan untuk secara manual mengidentifikasi sentimen di balik setiap ulasan. Oleh karena itu, analisis sentimen otomatis menjadi sangat krusial (Cahyaningtyas et al., 2021). Dengan memanfaatkan teknik *scraping* dan algoritma *natural language processing* (NLP), perusahaan dapat mengumpulkan dan menganalisis sejumlah besar ulasan dalam waktu singkat. Informasi yang diperoleh dari analisis ini, seperti proporsi ulasan positif dan negatif, serta topik yang paling sering dibicarakan, dapat memberikan wawasan berharga bagi perusahaan untuk meningkatkan kualitas produk atau layanan, memperbaiki strategi pemasaran, dan memahami lebih dalam kebutuhan pelanggan (Pasaribu & Sriani, 2023).

Algoritma Naïve Bayes, meskipun memiliki asumsi yang sederhana tentang independensi antar atribut, telah terbukti efektif dalam berbagai aplikasi, termasuk analisis sentimen. Kemampuannya dalam bekerja dengan data pelatihan yang terbatas menjadikannya pilihan yang menarik dalam

banyak kasus (Idris et al., 2023). Dalam konteks analisis sentimen, Python, dengan ekosistem perpustakaan yang kaya, menjadi bahasa pemrograman yang ideal untuk mengimplementasikan model Naïve Bayes. Prosesnya dimulai dari pra-pemrosesan data teks, seperti tokenisasi dan penghilangan stop words, hingga tahap klasifikasi sentimen. Dengan memanfaatkan kekuatan Python dan Naïve Bayes, kita dapat secara efisien menganalisis sejumlah besar ulasan dan mengkategorikannya menjadi sentimen positif atau negatif (Simanjuntak et al., 2023).

B. METODE PENELITIAN

Tahapan penelitian yang dilakukan pada penelitian ini menggunakan algoritma Naïve Bayes sebagai model dalam klasifikasi yang kemudian dilakukan proses evaluasi. Penjelasan secara detail mengenai tahapan metode penelitian disajikan pada Gambar 1 :



Gambar 1. Metodologi Penelitian

Studi Literatur

Penelitian ini telah dilakukan dengan mengumpulkan informasi dan data dari berbagai sumber pustaka, seperti buku referensi, artikel ilmiah, dan jurnal penelitian sebelumnya yang relevan dengan topik yang diteliti. Langkah ini bertujuan untuk membangun landasan teori yang kuat serta

mengidentifikasi penelitian serupa yang telah dilakukan sebelumnya.

Web Scraping

Pengumpulan data ulasan pengguna aplikasi Shopee dari Google Play Store dilakukan secara otomatis menggunakan teknik *web scraping* dengan bahasa pemrograman Python (Rahman et al., 2021). Proses ini berhasil mengumpulkan sebanyak 2000 data ulasan yang akan digunakan sebagai bahan analisis lebih lanjut.

Cleaning Data

Pembersihan data merupakan proses krusial dalam pengelolaan data yang bertujuan untuk meningkatkan kualitas data. Proses ini melibatkan identifikasi dan perbaikan atau penghapusan data yang mengandung kesalahan, ketidaklengkapan, ketidakkonsistenan, atau informasi yang tidak relevan. Dengan demikian, data yang telah dibersihkan dapat diandalkan untuk digunakan dalam berbagai analisis data, pelaporan, atau pengembangan model, sehingga menghasilkan hasil yang lebih akurat dan relevan.

Handing Missing Value-Ignore Tuple

Handing Missing Value-Ignore Tuple salah satu pendekatan dalam menangani data yang hilang (*missing values*) dalam sebuah dataset. Pendekatan ini dilakukan dengan mengabaikan atau menghapus seluruh baris (*tuple*) yang memiliki nilai hilang.

Text PreProcessing

Data mining melibatkan tahapan awal yang disebut *preprocessing*, di mana data mentah yang belum terstruktur diubah menjadi format yang lebih terorganisir dan mudah dipahami oleh komputer. Tujuan utama *preprocessing* adalah untuk mengatasi masalah seperti data yang berulang, data yang tidak relevan (*noise*), dan data yang kosong. Dalam konteks teks, *preprocessing* melibatkan langkah-langkah khusus untuk menyiapkan data teks sebelum dilakukan

analisis lebih lanjut.

1. *Case Folding*

Case folding merupakan suatu proses yang dilakukan untuk merubah semua huruf menjadi huruf kecil (*lower case*).

2. *Stopword*

Stopword berfungsi untuk menghapus kata yang tidak bermakna atau tidak berarti.

3. *Tokenizing*

Proses tokenisasi merupakan langkah awal dalam analisis teks yang melibatkan pembagian kalimat menjadi satuan-satuan kata yang lebih kecil. Proses ini dilakukan dengan memisahkan kata-kata berdasarkan spasi dan menghilangkan tanda baca yang tidak relevan. Tujuannya adalah untuk mendapatkan representasi teks yang lebih sederhana dan mudah diproses oleh komputer.

4. *Stemming*

Proses stemming merupakan proses menghilangkan kata dengan pengejaan yang buruk dengan menemukan kata dasar dan menghilangkan imbuhan. Setiap bahasa memiliki algoritma stemming unik. Stemming juga dilakukan untuk meningkatkan keseragaman data dan mengurangi daftar kata pada data latih.

Pemberian Label Pada Data

Setelah data terkumpul, langkah selanjutnya adalah memberikan label pada data. Label ini mencakup klasifikasi sentimen menjadi kategori positif dan negatif. Proses pemberian label ini bertujuan untuk mengidentifikasi sentimen positif dan negatif dari dataset yang telah dikumpulkan.

TF-IDF

Tujuan dari pemberian bobot menggunakan TF-IDF adalah untuk mengidentifikasi kata-kata yang paling relevan dan membedakan satu dokumen dengan dokumen lainnya. Dengan demikian, kata-kata yang dianggap penting akan lebih berpengaruh dalam proses analisis sentimen.

WordCloud

Word cloud atau awan kata adalah visualisasi data yang mengubah teks menjadi representasi grafis yang menarik. Dalam word cloud, kata-kata dalam teks disajikan dalam bentuk ukuran yang berbeda-beda. Ukuran setiap kata mencerminkan frekuensi kemunculannya dalam teks tersebut. Kata yang paling sering muncul akan ditampilkan dengan ukuran yang paling besar, sementara kata yang jarang muncul akan ditampilkan dengan ukuran yang lebih kecil. Hal ini memungkinkan kita untuk dengan cepat mengidentifikasi tema utama dan kata kunci dalam sebuah teks, tanpa harus membaca seluruh teks secara detail.

Klasifikasi Naïve Bayes

Naïve Bayes merupakan algoritma klasifikasi yang sering digunakan dalam pemrosesan bahasa alami. Algoritma ini bekerja berdasarkan asumsi independensi antara fitur-fitur (kata-kata) dalam sebuah dokumen. Dengan kata lain, kemunculan suatu kata dalam sebuah dokumen tidak dipengaruhi oleh kemunculan kata lainnya. Oleh karena itu, hasil klasifikasi suatu dokumen hanya ditentukan oleh informasi yang terkandung dalam dokumen tersebut tanpa memperhitungkan hubungan antar dokumen.

Evaluasi

Evaluasi dilaksanakan guna menilai kinerja model klasifikasi yang sudah dibangun. *Confusion matrix* seperti *accuracy*, *precision*, *recall*, dan *F1-score*, digunakan untuk mengukur sejauh mana model mampu mengklasifikasikan sentimen dengan akurat. Dari hasil proses klasifikasi pada *confusion matrix*, dihitung tiap nilai dengan *accuracy*, *precision*, *recall*, dan *F-1 score*.

Visualisasi

Setelah melakukan evaluasi dan klasifikasi sentimen, hasil penelitian akan

disajikan dalam bentuk visualisasi. Grafik dan diagram akan digunakan untuk menggambarkan distribusi sentimen positif, negatif, dan netral dalam data ulasan. Visualisasi ini memungkinkan kita untuk mengidentifikasi tren dan pola sentimen yang mendasari data, serta memberikan pemahaman yang lebih mendalam mengenai persepsi pengguna Shopee terhadap aplikasi tersebut.

C. HASIL DAN PEMBAHASAN

Web Scraping

Penelitian ini menggunakan teknik *web scraping* untuk mengumpulkan data ulasan pengguna Shopee dari Google Play Store. Memanfaatkan platform Google Colaboratory dan bahasa pemrograman Python, serta library *google-play-scraper* untuk mengotomatiskan proses pengambilan data. Dengan menggunakan ID unik aplikasi Shopee, kami berhasil mengumpulkan 2.000 ulasan terbaru pada tahun 2024. Data yang diperoleh kemudian disusun dalam format Excel untuk analisis lebih lanjut.

```
df_busu[['userName', 'score','at', 'content']].head()
```

Tabel 1. Hasil Scraping

Index	User Name	Score	at	content
1144	SopK elinci	1	9/10 /2024 1:03:22 PM	Udah bener benerr shopee kaya dulu gda biaya apapun kcuali proteksi keamanan, krna bisa saingan sma penjual offline,eh skrg malah nambah buat masalah, biaya layanan ditambah biaya penanganan, lucu !!
20	Baba n	5	9/10 /2024 12:48:47 PM	Manfaat sekali, saya edit lg ulasan thn 2021 ke thn skrg 2024 masih setia berbelanja online di aplikasi

Index	User Name	Score	at	content
				shope, karna berbelanja di sini tidak bikin saya bingung untuk memilih barang yang ingin saya beli, dari sistem cod maupun via bank, atau semacam nya.

Pelabelan

Data yang sudah diproses pra-pemrosesan kemudian diberi label, dengan label ini merujuk pada kamus opini positif dan negatif. Pelabelan ini akan memberikan informasi tentang sentimen atau opini dari setiap komentar dalam dataset. Tabel 2 menampilkan data yang sudah diberi label pada dataset tersebut.

```
def pelabelan(score):
    if score < 3:
        return 'Negatif'
    elif score == 4 :
        return 'Positif'
    elif score == 5 :
        return 'Positif'
my_df['Label'] = my_df
['score'].apply(pelabelan)
my_df.head(50)
```

Tabel 2. Pelabelan

index	content	score	label
998	apk ga guna punya saya tidak bisa cod lagi karna keterlambatan pengiriman padahal saya udah lama nunggu paket datang tapi yang bermasalah dari kurir nya ga ada nge Chet gak ada ngantar ujungnya kembali ke penjual	1	Negatif
189	habis di perbarui, tampilannya jadi berubah & ga bisa milih ekspedisi lagi klo checkout, auto pake shopee ekspres. mana lemot pula ini shopee	3	

index	content	score	label
	ekspres. kembalikan lagi dong pengaturan seperti semula. bebas pilih ekspedisi		

Cleaning Data

Proses ini bertujuan untuk menghilangkan tanda baca dari dataset. Hasil dari proses pembersihan data pada dataset ditampilkan dalam Tabel 3.

Tabel 3. Cleaning

index	content	Score	Label
8	Sangat mengecewakan, ternyata seperti itu kebijakan pengembalian barang dishopee sangat tidak bersahabat dengan pembeli padahal akunku sudah platinum. Mulai saat ini fix ini aplikasi nggak akan aku pakai lagi buat belanja	1	Negatif
9	satu hal yang aku tak suka di aplikasi shopee, tak bisa memperbesar gambar produk yang ingin kita beli dan itu sebenarnya sangat menyebalkan. serta tak ada tanda yg mudah di lihat konsumen untuk membedakan produk asli atau palsu	3	NaN

Handing Missing Value-Ignore Tuple

Handing Missing Value-Ignore Tuple salah satu pendekatan dalam menangani data yang hilang (*missing values*) dalam sebuah dataset.

Tabel 4. *Handing Missing Value*

index	content	Score	Label
17	Barangx bagus sekali dan saya sangat senang dan suka belanja di toko ini ...apalagi dg aplikasi shopee yg sangat membantu para pembeli...sukses	5	Positif

index	content	Score	Label
	terus buat shopeee		
21	Saya sudah jenuh pakai shopee..gk ada untungnya sya belanja terus sampai memmber gold dan hmpir masuk platinum..dan masalh pengiriman pake spx sekarnng lmbat..dan mau ngajukan paylater selalu di tolak..jdi saatnya saya beralih	1	Negatif

Text PreProcessing

Preprocessing merupakan tahapan awal yang sangat krusial dalam proses klasifikasi dokumen. Tahapan ini melibatkan serangkaian kegiatan yang bertujuan untuk mempersiapkan data teks agar siap diolah oleh algoritma klasifikasi.

1. Case Folding

Case folding dilakukan untuk menyederhanakan representasi teks dan mengurangi variasi dalam data. Dengan menghilangkan perbedaan antara huruf besar dan kecil serta tanda baca, kita dapat meningkatkan akurasi dalam proses analisis teks selanjutnya.

Tabel 5. *Case Folding*

index	content	score	label	Text_clean
2	Udah benerr shopee kaya dulu gda biaya apapun kcuali proteksi keamanan,krna bisa saingan sma penjual offline,eh skrg malah	1	Negatif	udah benerr shopee kaya dulu gda biaya apapun kcuali proteksi keamanan krna bisa

index	content	score	label	Text_clean
	nambah buat masalah ,biaya layanan ditambah biaya penanganan,lucu !! 😊😊			saingan sma penjual offlineeh skrg malah nambah buat masalah biaya layanan ditambah biaya penanganan lucu
4	Manfaat sekali, saya edit lg ulasan thn 2021 ke thn skrg 2024 masih setia berbelanja online di aplikasi shope, karna berbelanja di sini tidak bikin saya bingung untuk memilih barang yang ingin saya beli, dari sistem cod maupun via bank, atau semacam nya.. 🙏	5	Positif	manfaat sekali saya edit lg ulasan thn ke thn skrg masih setia berbelanja online di aplikasi shope karna berbelanja di sini tidak bikin saya bingung untuk memilih barang yang ingin saya beli dari sistem cod maupun via bank atau semacamnya

2. Stopword

Langkah ini dimaksudkan untuk menghapus kata-kata umum yang sering dianggap tidak memberikan informasi penting dalam analisis teks, seperti "yang", "dan", "dengan", dan sejenisnya. Dalam penelitian ini, penghapusan stopwords

dilakukan dengan memanfaatkan *corpus* atau daftar kata-kata berbahasa Indonesia yang biasanya dianggap tidak relevan dalam pemrosesan teks. Hasil dari proses penghapusan stopword pada dataset ditampilkan dalam Tabel 6.

Tabel 6. Stopword

inde x	content	score	label	Text_stopword
6	Ya enak aja jdi klo di shope pengalamn saya blm pernah ketipu barang yg dikirim sesuai orderan, klo seandainya ada masalah penipuan/nakal ada laporan ke pihak shope buru ² di blokir aja	5	Positif	ya enak aja jdi klo shope pengalamn saya blm ketipu barang yg dikirim sesuai orderan klo seandainya penipuan nakal laporan shope buru blokir aja
7	Sangat mengecewakan, ternyata seperti itu kebijakan pengembalian barang dishopee sangat tidak bersahabat dengan pembeli padahal akunku sudah platinum. Mulai saat ini fix ini aplikasi nggak akan aku pakai lagi buat belanja	1	Negatif	mengecewakan kebijakan pengembalian barang dishopee pembeli akunku platinum fix aplikasi nggak pakai belanja

3. Tokenizing

Proses Ini adalah proses yang digunakan untuk memotong suatu kalimat menjadi kata-kata yang terpisah berdasarkan spasi. Berikut contoh hasil dari proses tokenizing :

Tabel 7. Tokenizing

inde x	content	score	label	Text_tokenizing
1	apk ga guna punya saya tidak bisa cod lagi karna keterlambatan pengiriman padahal saya udah lama nunggu paket datang tapi yang bermasalah dari kurir nya ga ada nge Chet gak ada ngantar ujungnya kembali ke penjual		Negatif	apk,ga,cod,karna,keterlambatan,pengiriman,udah,nunggu,paket,bermasalah,kurir,nya,ga,nge,chet,gak,ngantar,ujungnya,pembelian
2	Manfaat sekali, saya edit lg ulasan thn 2021 ke thn skrg 2024 masih setia berbelanja online di aplikasi shopee, karna berbelanja di sini tidak bikin saya bingung untuk memilih barang yang ingin saya beli, dari sistem cod maupun via bank, atau semacam nya..		Positif	manfaat,edit,lg,ulasan,thn,thn,skrg,setia,berbelanja,online,aplikasi,shopee,karna,berbelanja,di,sini,tidak,bikin,saya,bingung,untuk,memilih,barang,yang,ingin,saya,beli,dari,sistem,cod,maupun,via,bank,atau,semacam,nya..

4. Stemming

Stemming dengan Sastrawi dilakukan untuk mengurangi variasi bentuk kata dalam teks bahasa Indonesia. Dengan menghilangkan imbuhan, kita dapat mengelompokkan kata-kata yang memiliki arti serupa menjadi satu kelompok. Hal ini sangat berguna dalam proses analisis sentimen atau pengelompokan teks.

```
!pip install Sastrawi
from Sastrawi.Stemmer.StemmerFactory
import StemmerFactory
factory = StemmerFactory()
```

stemmer = factory.create_stemmer()

Tabel 8. *Stemming*

ind ex	content	scor e	labe l	Text_st emmin g
15	Barangx bagus sekali dan saya sangat senang dan suka belanja di toko ini ...apalagi dg aplikasi shopee yg sangat membantu para pembeli...suk ses terus buat shopee	5	Posi tif	barangx bagus senang suka belanja toko dg aplikasi shopee yg bantu pembeli sukses shopee

Split Data

Pembagian data menjadi data latih dan data uji merupakan langkah penting dalam pengembangan model analisis sentimen. Untuk yang pertama pembagian data 80:20 dan pembagian data yang kedua 60:40. Data latih akan digunakan oleh model untuk belajar mengenali pola sentimen, sedangkan data uji akan digunakan untuk mengukur kinerja model pada data yang belum pernah dilihat sebelumnya. Dengan menggunakan parameter `random_state`, kita dapat memastikan bahwa hasil evaluasi model dapat diperbandingkan antar percobaan.

```
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(data_clean['content'], data_clean['label'],
                                                test_size = 0.20,
                                                random_state = 0)

#membagi data menjadi data training dan testing dengan test_size = 0.40 dan random state nya 0
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(data_clean['content'], data_clean['label'],
                                                test_size = 0.40, random_state = 0)
```

Gambar 2. Split Data

TF-IDF

Dalam penelitian ini, telah dilakukan analisis mendalam terhadap sentimen

pengguna aplikasi Shopee dengan menggunakan teknik pengumpulan data secara otomatis (*web scraping*) dan model klasifikasi Naive Bayes. Dari 2000 ulasan yang dikumpulkan, berhasil diproses 1739 ulasan untuk dianalisis. Hasil analisis menunjukkan bahwa mayoritas pengguna memiliki sentimen negatif terhadap aplikasi Shopee. Meskipun model ini mampu mengklasifikasikan ulasan dengan akurasi yang cukup baik (83%), masih terdapat ruang untuk perbaikan, terutama dalam mengidentifikasi ulasan positif.

Penggunaan `TfidfVectorizer` memungkinkan kita untuk membuat representasi vektor untuk teks di mana tiap elemen vektor mencerminkan frekuensi kata dalam dokumen. Langkah awal setelah melakukan impor library yang diperlukan adalah membuat objek `TfidfVectorizer` dan kemudian melatihnya pada data *train* menggunakan metode `fit()` seperti hasil pada gambar 4.10 dimana proses ini berguna untuk mempelajari korpus dari data latih, yang akan digunakan nantinya untuk mengonversi teks menjadi vektor.

```
[183] from sklearn.feature_extraction.text import TfidfVectorizer

tfidf_vectorizer = TfidfVectorizer()
tfidf_train = tfidf_vectorizer.fit_transform(X_train)
tfidf_test = tfidf_vectorizer.transform(X_test)

print(X_train.shape)
print(y_train.shape)
print(X_test.shape)
print(y_test.shape)

(1392,)
(1392,)
(348,)
(348,)

[69] from sklearn.feature_extraction.text import TfidfVectorizer

tfidf_vectorizer = TfidfVectorizer()
tfidf_train = tfidf_vectorizer.fit_transform(X_train)
tfidf_test = tfidf_vectorizer.transform(X_test)

print(X_train.shape)
print(y_train.shape)
print(X_test.shape)
print(y_test.shape)

(1043,)
(1043,)
(696,)
(696,)
```

Gambar 3. TF-IDF

Word Cloud

$$\frac{TF + TN}{TP + FP + TN + FN}$$

$$\frac{208 + 80}{208 + 50 + 80 + 11}$$

$$\frac{288}{345}$$

$$= 0,83$$

b. *Precision*

$$\frac{TP}{TP + FP}$$

$$\frac{208}{208 + 50}$$

$$\frac{208}{258}$$

$$= 0,81$$

c. *Recall*

$$\frac{TP}{TP + FN}$$

$$\frac{208}{208 + 11}$$

$$\frac{208}{219}$$

$$= 0,95$$

d. *F1-Score*

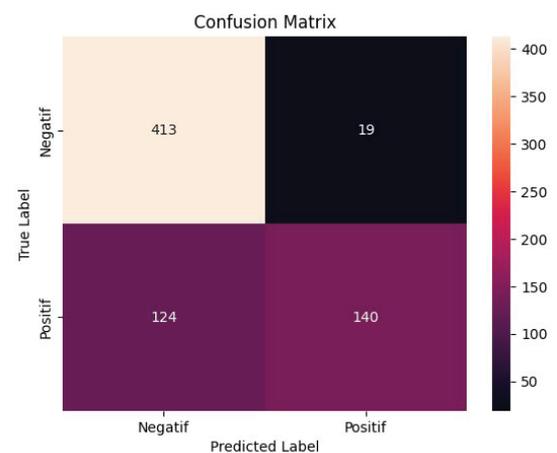
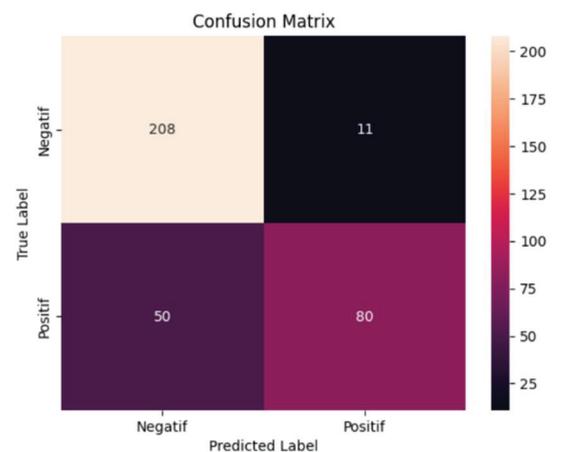
$$2x \frac{P \times R}{P + R}$$

$$2x \frac{0,81 \times 0,95}{0,81 + 0,95}$$

$$2x \frac{0,7695}{1,76}$$

$$= 0,87$$

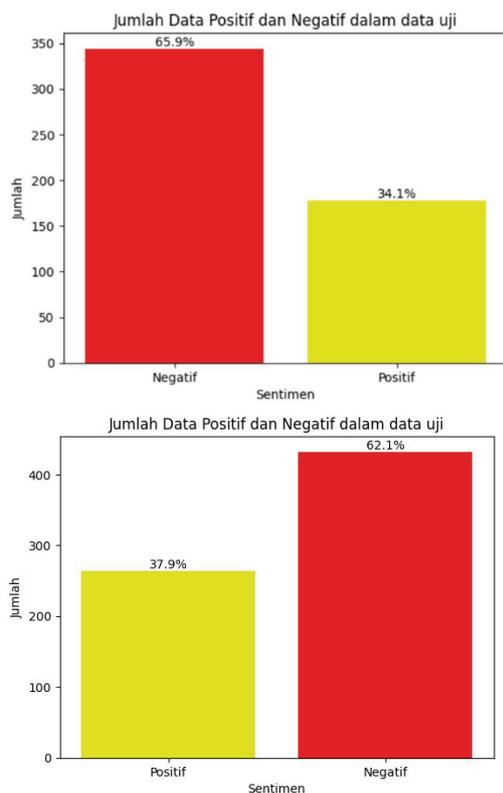
Confusion matrix divisualisasikan sebagai representasi grafis dari 118egat yang digunakan untuk mengevaluasi kinerja model klasifikasi. Pada sumbu x (horizontal) kelas yang diprediksi oleh model ditampilkan, sedangkan pada sumbu y (negatif) kelas sebenarnya dari data uji ditampilkan. Visualisasi ini menggunakan *heatmap* (peta panas) dari *library seaborn (sns.heatmap)*. Seperti pada gambar 6 *heatmap* digunakan untuk memvisualisasikan data dalam bentuk matriks dengan menggunakan skala warna untuk menggambarkan nilai di setiap sel. Tujuan dari visualisasi ini adalah untuk mempermudah pemahaman tentang bagaimana model mengklasifikasikan data uji ke dalam kelas negatif dan positif, serta seberapa baik model tersebut dalam memprediksi kelas yang benar.



Gambar 6. *Confusion Matrix*

Visualisasi

Pada tahap ini, dilakukan proses visualisasi untuk menggambarkan hasil evaluasi dan klasifikasi data secara grafis. Tujuannya adalah untuk memberikan pemahaman tentang bagaimana tanggapan pengguna shopee terhadap ulasan di google *Play Store* serta menyajikan gambaran tentang bagaimana aplikasi diberi ulasan negatif atau positif oleh pengguna. Dengan membedakan antara sentimen positif dan negatif, kita dapat mengidentifikasi aspek-aspek yang dianggap baik atau buruk oleh pengguna shopee. Visualisasi ini memanfaatkan *library Python seaborn* untuk membuat plot menggunakan metode *countplot*, yang menampilkan jumlah data dalam kategori yang berbeda, serta menggunakan *matplotlib* untuk menampilkan hasil visualisasi. Dari gambar 7 terlihat bahwa jumlah sentimen negatif lebih mendominasi dibandingkan dengan sentimen positif.



Gambar 7. Proses Visualisasi

D. PENUTUP

Berdasarkan hasil dari pembahasan disertai hasil analisis yang telah dipaparkan menghasilkan kesimpulan diantaranya didapat cara untuk menganalisa ulasan shopee pada *play store* adalah menggunakan metode Naïve Bayes. Untuk proses pengumpulan data ulasan dilakukan dengan *scrapping* pada Google Play Store menggunakan python, menghasilkan data mentah sebanyak 2000 data ulasan. Lalu data mentah tersebut diproses menggunakan python sehingga menghasilkan data bersih sebanyak 1739 data.

Performa algoritma Naïve Bayes menghasilkan akurasi yaitu 83%. Untuk nilai positif *precision* 88%, *recall* 62%, dan *f1 score* 72%. Sedangkan untuk akurasi negatif nilai yaitu 81% *precision*, 95% *recall*, dan *f1 score* 87%. Dari hasil ini dapat disimpulkan bahwa mayoritas sentimen pengguna ulasan shopee di Google Play Store lebih cenderung negatif.

E. DAFTAR PUSTAKA

- Agustina, N., Citra, D. H., Purnama, W., Nisa, C., & Kurnia, A. R. (2022). Implementasi Algoritma Naive Bayes untuk Analisis Sentimen Ulasan Shopee pada Google Play Store. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, 2(1), 47–54. <https://doi.org/10.57152/malcom.v2i1.195>
- Asih, E. M. (2024). Analisis pada Shopee Sebagai E-Commerce Terpopuler di Indonesia. *Jurnal Ekonomi Bisnis Antartika*, 2(1), 73–79. <https://doi.org/10.70052/jeba.v2i1.299>
- Cahyaningtyas, C., Nataliani, Y., & Widiyari, I. R. (2021). Analisis Sentimen Pada Rating Aplikasi Shopee Menggunakan Metode Decision Tree Berbasis SMOTE. *AITI: Jurnal Teknologi Informasi*, 18(2), 173–184.

- <https://doi.org/10.24246/aiti.v18i2.173-184>
- Idris, I. S. K., Mustofa, Y. A., & Salihi, I. A. (2023). Analisis Sentimen Terhadap Penggunaan Aplikasi Shopee Menggunakan Algoritma Support Vector Machine (SVM). *Jambura Journal of Electrical and Electronics Engineering*, 5(1), 32–35. <https://doi.org/10.37905/jjee.v5i1.16830>
- Nurian, A., Ma'arif, M. S., Amalia, I. N., & Rozikin, C. (2024). Analisis Sentimen Pengguna Aplikasi Shopee Pada Situs Google Play Menggunakan Naive Bayes Classifier. *JITET: Jurnal Informatika Dan Teknik Elektro Terapan*, 12(1), 97–105. <https://doi.org/10.23960/jitet.v12i1.3631>
- Oktia, S., Soesanto, S., Purwanto, H., & Rochendi, T. (2024). Pengaruh Harga, Kualitas dan Jenis Produk Terhadap Pendapatan Toko Online P.T. Clothme Jakarta di Aplikasi Shopee. *REMITTANCE: Jurnal Akuntansi Keuangan Dan Perbankan*, 5(2), 71–82. <https://doi.org/10.56486/remittance.vol5no2.647>
- Pasaribu, N. A., & Sriani. (2023). The Shopee Application User Reviews Sentiment Analysis Employing Naïve Bayes Algorithm. *IJSECS: International Journal Software Engineering and Computer Science*, 3(3), 194–204. <https://doi.org/10.35870/ijsecs.v3i3.1699>
- Rahman, A., Utami, E., & Sudarmawan. (2021). Sentimen Analisis Terhadap Aplikasi pada Google Playstore Menggunakan Algoritma Naïve Bayes dan Algoritma Genetika. *Komtika: Jurnal Komputasi Dan Informatika*, 5(1), 60–71. <https://doi.org/10.31603/komtika.v5i1.5188>
- Simanjuntak, R. L., Siagian, T. R., Anggriani, V., & Arnita, A. (2023). Analisis Sentimen Ulasan Pada Aplikasi E-Commerce Shopee Dengan Menggunakan Algoritma Naïve Bayes. *Teknik: Jurnal Teknik Mesin, Elektro Dan Ilmu Komputer*, 3(3), 23–39. <https://doi.org/10.55606/teknik.v3i3.2411>