

Application of Multinomial Naïve Bayes for Sentiment Classification on Bukalapak Reviews

Musyafa Faeang Ogya Widi^{1*}, Dona Yuliawati^{2**}

^{*}Sistem Informasi, Institut Informatika Dan Bisnis Darmajaya
farandogya@gmail.com¹, donayuliawati@darmajaya.ac.id²

Article Info

Article history:

Received 2025-11-01

Revised 2025-11-29

Accepted 2025-12-10

Keyword:

*Sentiment Analysis,
Multinomial Naïve Bayes,
E-Commerce,
Bukalapak,
Customer Reviews.*

ABSTRACT

This study investigates sentiment analysis on user reviews from Bukalapak, a major Indonesian e-commerce platform, using the Multinomial Naïve Bayes (MNB) classifier. The study focuses on tackling the challenge of data imbalance and the linguistic complexities of Indonesian, such as slang, affixes, and negation, which are common in user reviews. Data was collected through web scraping from Bukalapak's app on the Google Play Store, resulting in a dataset of 19,999 reviews. A structured preprocessing pipeline was employed, including text normalization, tokenization, stopword removal, stemming, and term frequency-inverse document frequency (TF-IDF) weighting to prepare the data. The sentiment analysis results show that the model performs well in categorizing neutral reviews (accuracy 81%), but struggles with positive and negative sentiments due to data imbalance, leading to lower accuracy for these categories. The study highlights the effectiveness of Multinomial Naïve Bayes in large-scale sentiment analysis tasks in the e-commerce domain, particularly for platforms with large volumes of user-generated content. The study also introduces SMOTE (Synthetic Minority Over-sampling Technique) for handling data imbalance and k-fold cross-validation for model evaluation, significantly improving the model's reliability. The research concludes that sentiment analysis can greatly benefit e-commerce platforms by improving customer service, informing product management decisions, and providing valuable insights for business strategies.



This is an open access article under the [CC-BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.

I. PENDAHULUAN

Pertumbuhan pengguna internet di Indonesia terus mengalami peningkatan yang signifikan setiap tahunnya. Indonesia kini menduduki peringkat keempat dunia setelah China, India, dan Amerika Serikat dalam hal pertumbuhan jumlah pengguna internet. Berdasarkan laporan dari Asosiasi Penyelenggara Jasa Internet Indonesia (APJII), penetrasi internet di Indonesia telah mencapai 77,02% pada tahun 2021-2022, meningkat dari 64,8% pada 2018 dan 73,7% pada 2019-2020. Tren peningkatan ini menunjukkan potensi besar bagi perkembangan ekonomi digital di Indonesia, salah satunya melalui sektor e-commerce [1].

Salah satu platform e-commerce terbesar di Indonesia adalah Bukalapak. Bukalapak, yang didirikan pada tahun 2010 oleh Achmad Zaky, dkk, telah berkembang pesat

menjadi salah satu unicorn di Indonesia. Pada tahun 2017, Bukalapak mencatatkan namanya dalam jajaran startup unicorn Indonesia dengan jumlah kunjungan yang menempatkannya di posisi ketiga di Indonesia setelah platform e-commerce lainnya. Bukalapak awalnya berfokus pada menyediakan toko daring untuk Usaha Kecil dan Menengah (UKM), namun kini telah memperluas layanannya dengan berbagai fitur, termasuk marketplace, teknologi finansial, dan layanan logistik, untuk memenuhi berbagai kebutuhan pelanggan [2], [3].

Salah satu fitur penting yang dimiliki oleh Bukalapak adalah sistem ulasan produk yang memungkinkan pengguna untuk memberikan penilaian dan komentar terkait produk yang telah dibeli. Ulasan ini, yang terdiri dari peringkat bintang dan komentar teks, membantu calon pembeli untuk membuat keputusan pembelian. Namun, dengan jumlah

ulasan yang sangat banyak, pengelolaan dan analisis ulasan menjadi tantangan besar bagi platform tersebut. Menurut Aswin dan Wahidun (2016), perusahaan yang cerdas harus memahami proses pengambilan keputusan pelanggan, termasuk ulasan dan penilaian yang mereka berikan terhadap produk atau layanan [4].

Untuk mengatasi tantangan ini, analisis sentimen terhadap ulasan pengguna dapat membantu dalam memahami pendapat konsumen, baik yang positif, negatif, maupun netral. Dengan menggunakan metode otomatis seperti Multinomial Naïve Bayes, Bukalapak dapat lebih cepat mengklasifikasikan ulasan produk dan memprediksi apakah suatu produk pantas direkomendasikan. Mengingat volume ulasan yang sangat besar, pengelolaan ulasan manual menjadi sangat tidak efisien. Oleh karena itu, penerapan model Multinomial Naïve Bayes untuk otomatisasi klasifikasi sentimen menawarkan solusi yang efektif dan efisien dalam memproses data dalam jumlah besar [5].

Salah satu metode yang dapat digunakan adalah Multinomial Naïve Bayes, yang telah terbukti efektif dalam melakukan klasifikasi teks, termasuk untuk analisis sentimen. Metode ini memanfaatkan model probabilistik untuk mengklasifikasikan ulasan menjadi sentimen positif, negatif, atau netral, serta efisien dalam menangani volume data yang besar, yang sangat relevan untuk aplikasi e-commerce [5], [6].

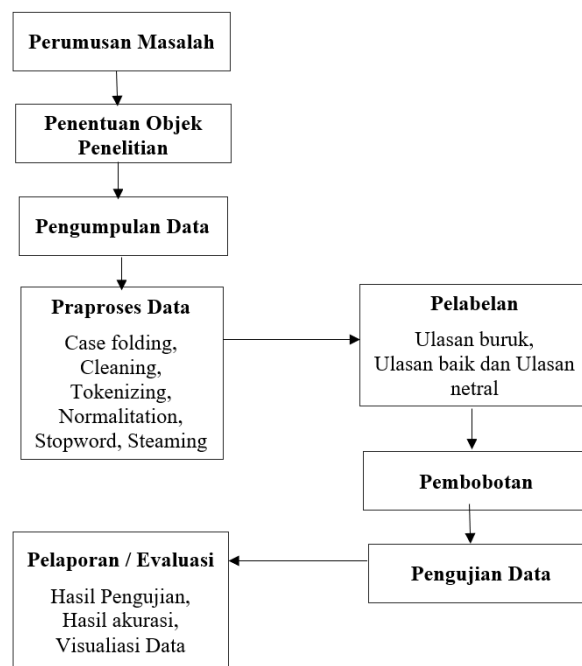
Penelitian ini bertujuan untuk menerapkan metode Multinomial Naïve Bayes pada analisis sentimen ulasan produk di Bukalapak, sebuah platform e-commerce besar di Indonesia, yang sebelumnya belum banyak dijelajahi dalam literatur terkait [7]. Fokus utama penelitian ini adalah mengatasi ketidakseimbangan data yang sering muncul dalam analisis sentimen, serta meningkatkan akurasi model dalam mengklasifikasikan sentimen positif, negatif, dan netral pada ulasan pengguna. Dengan membandingkan kinerja Multinomial Naïve Bayes dengan model lain, penelitian ini diharapkan dapat memberikan wawasan lebih dalam mengenai efektivitas metode ini dalam konteks analisis sentimen Bahasa Indonesia, khususnya untuk e-commerce.

Selain itu, penelitian ini juga bertujuan untuk memberikan kontribusi praktis kepada Bukalapak dalam memahami sentimen pelanggan, yang dapat digunakan untuk meningkatkan layanan pelanggan, pengelolaan produk, serta pengambilan keputusan bisnis yang lebih tepat. Dengan adanya teknologi ini, Bukalapak dapat memanfaatkan ulasan pelanggan secara lebih efisien untuk memperbaiki kualitas produk dan layanan yang ditawarkan [8].

II. METODE

A. Alur Tahapan Penelitian

Proses analisis sentimen pada ulasan pengguna Bukalapak dilakukan melalui beberapa tahapan yang terstruktur. Setiap tahapan bertujuan untuk mempermudah dan memastikan bahwa hasil analisis dapat menggambarkan dengan akurat sentimen yang terkandung dalam ulasan pengguna. Tahapan penelitian tersebut adalah sebagai berikut.



Gambar 1. Alur Penelitian.

Proses analisis sentimen pada ulasan pengguna Bukalapak dilakukan menggunakan pendekatan Multinomial Naïve Bayes yang memanfaatkan teori-teori dalam pengolahan teks dan analisis sentimen [9][10]. Tahapan-tahapan dalam metode penelitian ini terdiri dari beberapa proses yang bertujuan untuk mempermudah dan memastikan bahwa hasil analisis dapat menggambarkan dengan akurat sentimen yang terkandung dalam ulasan pengguna.

B. Pengumpulan Data

Pada tahap awal, data yang digunakan dalam penelitian ini dikumpulkan melalui teknik scraping dari aplikasi Bukalapak yang ada di Google Play Store. Data yang dikumpulkan berupa ulasan pengguna yang terdiri dari rating dan komentar teks yang diberikan oleh pengguna produk. Pengumpulan data dilakukan dengan memanfaatkan teori Electronic Word of Mouth (eWOM), yang mengacu pada penggunaan ulasan dan rekomendasi dari konsumen untuk memahami pengaruh pengalaman pelanggan terhadap keputusan pembelian orang lain [11][12].

C. Alur Tahapan Penelitian

Preprocessing atau praproses data merupakan tahap yang sangat penting dalam analisis teks untuk mempersiapkan data menjadi bentuk yang siap untuk dianalisis. Tahapan preprocessing mencakup beberapa fungsi yang teoretis sebagai berikut.

1) Normalisasi Teks

Proses untuk mengubah teks yang tidak baku (slang atau kata tidak standar) menjadi bentuk yang sesuai dengan Kamus Besar Bahasa Indonesia (KBBI). Tujuan dari normalisasi

adalah agar teks dapat dikenali dengan baik dalam analisis selanjutnya. Normalisasi termasuk konversi kata yang tidak baku menjadi kata baku sesuai dengan kaidah bahasa resmi [13].

2) *Tokenisasi*.

Memisahkan teks menjadi token-token kata yang lebih kecil agar bisa diproses lebih lanjut. Langkah ini penting dalam text preprocessing untuk memudahkan analisis sentiment [14].

3) *Penghilangan Kata Tidak Penting (Stopword Removal)*

Menghapus kata-kata yang tidak memberikan kontribusi signifikan terhadap analisis sentimen, seperti kata hubung yang sering muncul dalam teks, sesuai dengan teori stopwords removal dalam pemrosesan teks [15].

4) *Stemming*

Mengubah kata turunan menjadi bentuk dasar atau akar kata. Stemming dilakukan menggunakan algoritma yang sudah diterapkan dalam berbagai penelitian di bidang analisis teks. Proses stemming bertujuan untuk mengubah kata turunan (misalnya, "berjalan", "berjalanlah") menjadi bentuk dasarnya ("jalan"). Hal ini dilakukan untuk mengurangi variasi kata dan membuat analisis lebih sederhana [16].

D. Pembobotan dengan TF-IDF

Setelah teks diproses, tahap selanjutnya adalah pemberian bobot pada setiap kata menggunakan metode Term Frequency-Inverse Document Frequency (TF-IDF). TF-IDF adalah metode pembobotan statistik yang digunakan untuk mengukur pentingnya sebuah kata dalam koleksi dokumen. Pembobotan TF-IDF digunakan untuk memastikan bahwa kata-kata yang sering muncul dalam satu dokumen tetapi jarang muncul dalam seluruh koleksi dokumen diberi bobot tinggi. Konsep ini menjadikan TF-IDF sangat efektif dalam mengidentifikasi kata-kata kunci dalam analisis sentiment [17].

$$TF - IDF(t, d) = TF(t, d) \times \log \left(\frac{N}{DF(t)} \right)$$

dimana $TF(t, d)$ adalah frekuensi kemunculan kata t dalam dokumen d , $DF(t)$ adalah jumlah dokumen yang mengandung kata t , dan N adalah jumlah total dokumen dalam koleksi.

E. Multinomial Naïve Bayes Classifier

Setelah data diberi pembobotan menggunakan TF-IDF, tahap selanjutnya adalah menerapkan Multinomial Naïve Bayes Classifier untuk mengklasifikasikan sentimen ulasan menjadi tiga kategori: positif, negatif, dan netral. Berikut adalah penjelasan teori dan fungsi teoretis dari Multinomial Naïve Bayes yang digunakan dalam analisis. Multinomial Naïve Bayes adalah algoritma klasifikasi probabilistik yang menghitung probabilitas kelas (sentimen) berdasarkan distribusi kata dalam dokumen. Dengan asumsi bahwa kata-kata dalam dokumen bersifat independen (asumsi Naïve),

model ini menghitung probabilitas setiap kelas dengan menggunakan rumus Teorema Bayes [18][19]:

$$P(C | X) = \frac{P(C) \cdot P(X | C)}{P(X)}$$

Probabilitas kelas $P(C | X)$ dihitung untuk setiap kelas C (positif, negatif, netral), berdasarkan fitur X yang merupakan kata-kata dalam ulasan. Prior probability $P(C)$ mengukur seberapa sering kelas C muncul dalam data pelatihan, sedangkan likelihood $P(X | C)$ mengukur probabilitas kata-kata X muncul dalam kelas C . Probabilitas total $P(X)$ dihitung berdasarkan semua kata dalam ulasan dan biasanya diabaikan dalam perhitungan klasifikasi karena konstan untuk semua kelas [21-24]. Multinomial Naïve Bayes mengasumsikan bahwa kata-kata dalam dokumen adalah independen, yang memungkinkan algoritma menghitung probabilitas setiap kata secara terpisah dan menggabungkannya untuk menentukan kelas dengan probabilitas tertinggi. Dalam penelitian ini, Naïve Bayes digunakan untuk mengklasifikasikan sentimen ulasan pengguna berdasarkan kata-kata yang terkandung dalam teks ulasan tersebut.

F. Evaluasi Model

Evaluasi model dilakukan dengan menggunakan metrik akurasi, presisi, recall, dan F1-score. Akurasi mengukur persentase prediksi yang benar, sedangkan presisi dan recall mengukur akurasi dalam mengklasifikasikan sentimen positif. F1-score memberikan rata-rata harmonis antara presisi dan recall. Semua metrik ini dihitung berdasarkan confusion matrix, yang menunjukkan True Positives (TP), True Negatives (TN), False Positives (FP), dan False Negatives (FN), untuk menilai kinerja model dalam klasifikasi sentimen [20][21].

III. HASIL DAN PEMBAHASAN

Bagian ini menyajikan hasil penerapan algoritma Multinomial Naïve Bayes untuk analisis sentimen ulasan pengguna Bukalapak. Pembahasan akan fokus pada evaluasi kinerja model, termasuk akurasi dan efektivitas dalam mengklasifikasikan sentimen positif, negatif, dan netral berdasarkan data ulasan yang telah diproses. Pengumpulan data dilakukan menggunakan google play scraper untuk menghubungkan ulasan dari aplikasi Bukalapak di Google Play Store. Data yang dikumpulkan disimpan dalam format CSV dan terdiri dari 19.999 ulasan yang akan diproses lebih lanjut dengan model Naïve Bayes. Berikut adalah kode singkat yang digunakan untuk mengumpulkan data.

Bahasa Indonesia memiliki tantangan linguistik yang unik dalam analisis sentimen, seperti penggunaan slang, imbuhan, negasi, dan ambiguitas kata yang sering ditemukan dalam ulasan produk. Misalnya, kata-kata slang seperti "gokil" (luar biasa) atau "santuy" (santai) dapat memiliki makna yang berbeda tergantung pada konteksnya. Selain itu, penggunaan imbuhan (misalnya, "berkendara" atau "menyukai") memperkenalkan variasi kata yang perlu distem untuk

menyederhanakan analisis. Negasi dalam kalimat seperti "tidak suka" atau "kurang baik" juga dapat membalikkan makna sentimen yang diekspresikan. Untuk mengatasi tantangan-tantangan ini, model ini mengimplementasikan normalisasi teks, tokenisasi, dan stemming untuk mengubah kata-kata tidak baku menjadi bentuk standar, serta mengenali pola negasi yang dapat mempengaruhi hasil klasifikasi sentimen.

```
from google_play_scraper import reviews, Sort

app_id = 'com.bukalapak.android'
reviews_data = []
for result, token in reviews(app_id, lang='id', country='id',
sort=Sort.MOST_RELEVANT, count=100):
    reviews_data.extend(result)
```

Untuk mengatasi masalah ketidakseimbangan data, penulis menggunakan SMOTE (Synthetic Minority Over-sampling Technique) untuk menyeimbangkan jumlah sentimen positif, negatif, dan netral dalam dataset. Selain itu, proses k-fold cross-validation diterapkan untuk menguji keandalan model dengan membagi dataset menjadi beberapa bagian dan memastikan bahwa model diuji dengan data yang berbeda pada setiap iterasi, yang dapat mengurangi risiko overfitting dan memberikan evaluasi kinerja yang lebih robust.

A. Hasil Preprocessing

Pada tahap preprocessing, data ulasan yang dikumpulkan melalui proses scraping dari Google Play Store diolah dengan beberapa tahapan untuk menyiapkan data agar siap digunakan dalam model Naïve Bayes.

Pada tahap case folding huruf kapital pada semua dokumen tweet diubah menjadi huruf kecil contoh tahap konversi huruf menjadi kecil dapat dilihat pada table berikut.

TABEL I
CASE FOLDING

Input	Output
@Yoghaaaa Hi, pesanan kamu sudah dalam pengiriman. Maksimal kamu menerima produk 16/02/2017. Mohon ditunggu ya. Thanks –Bella	@yoghaaaa hi, pesanan kamu sudah dalam pengiriman. maksimal kamu menerima produk 16/02/2017. mohon ditunggu ya. thanks –bella

Selanjutnya, dilakukan tahapan cleaning merupakan tahap pembersihan kata yang tidak berpengaruh sama sekali terhadap hasil klasifikasi sentimen.. Contoh dari atribut yang tidak penting tersebut adalah yaitu mention yang diawali dengan atribut ('@'), hastag yang diawali dengan atribut ('#'), link yang diawali dengan atribut ('http', 'bit.ly') dan karakter simbol (~!@#\$%^&* _+?<>.,?:{}[]|) contoh tahap menghapus simbol pada table berikut.

TABEL II
CLEANING

Input	Output
@Yoghaaaa Hi, pesanan kamu sudah dalam pengiriman. Maksimal kamu menerima produk 16/02/2017. Mohon ditunggu ya. Thanks –Bella	hi, pesanan kamu sudah dalam pengiriman maksimal kamu menerima produk 16/02/2017 mohon ditunggu ya. thanks –bella

Selanjutnya, pada tahap *normalization* ini dilakukan pengubahan kata yang tidak sesuai dengan dokumen kamus bahasa indonesia, sehingga dapat mengurangi hasil sentimen dokumen contoh tahap konversi inggris ke indonesia berikut.

TABEL III
NORMALIZATION

Input	Output
hi, bukalapak utk sementara tdk dptdiakses & masih dalam proses utk memulihkannya mohon maaf atas ketidaknyamanannya bl	hi bukalapak untuk sementara tidak dapat diakses amp masih dalam proses untuk memulihkannya mohon maaf atasketidaknyamanannya bl

Proses selanjutnya adalah tahap konversi kata baku, karena standar bahasa yang dipakai pada penelitian ini adalah bahasa Indonesia baku pada tabel berikut.

TABEL IV
CONTOH TAHAP KONVERSI KATA BAKU (NORMALIZATION)

Input	Output
hai yunni mohon maaf mohon informasikan terlebih dahulu email akun tokopedia via dm	hai yunni mohon maaf mohon informasi lebih dahulu email akun via

Tahap yang terakhir dalam proses Normalization adalah mengubah kata inggris ke Indonesia, berikut contoh konversi kata inggris pada tabel.

TABEL V
CONTOH TAHAP KONVERSI KATA INGGRIS (NORMALIZATION)

Input	Output
ulasan customer service yang menggunakan gosend hari ini thanksbuat bintang nya	ulasan customer service yang menggunakan gosend hari ini terimakasih buat bintang nya

Tahap *tokenizing* Kata dalam dokumen yang dimaksud adalah kata yang dipisah oleh spasi. Sehingga hasil dari proses ini merupakan kata tunggal yang dimasukkan ke dalam *database* untuk keperluan pembobotan.

TABEL VI
TOKENIZING

Input	Output
pesanan sudah dalam pengiriman, jadi sudah tidak bisa dibatalkan	'pesanan' 'sudah' 'dalam' 'pengiriman' 'jadi' 'sudah' 'tidak' 'bisa' 'dibatalkan'

Tahap *stopword removal* merupakan tahap menghilangkan kata yang tidak sesuai dengan topik dokumen, jika ada kata tersebut tidak mempengaruhi akurasi dalam klasifikasi sentimen dokumen.

TABEL VII
STOPWORD

Input	Output
bukan kah resinya sudah saya kirim dan juga sudah di konfirmasi oleh Bukalapak klo barangnya sudah sampai di gudang Bukalapak	bukan resinya kirim Konfirmasi Bukalapak klo barangnya Gudang Bukalapak

Stemming Jika ditemukan maka diasumsikan bahwa kata tersebut adalah root word. Maka metode berhenti. Inflection Suffixes (“-lah”, “-kah”, “-ku”, “-mu”, atau “-nya”) dibuang. Jika berupa particles (“-lah”, “-kah”, “-tah” atau “-pun”).

TABEL VII
Stemming

Input	Output
kalau tetap tidak terkirim pakatnya dianggap batal gimana kalau batal pesanan min pembayaran via cod	kalau tetap tidak kirim paket anggap batal gimana pesan kalau batal pesan min bayar via cod

B. Hasil Pelabelan Polaritas Dan Subjectivity

Pelabelan sentimen menggunakan dua faktor utama, yaitu polaritas dan subjektivitas. Polaritas menentukan apakah teks mengungkapkan sentimen positif, negatif, atau netral dengan

membandingkan kata-kata dalam teks dengan daftar kata positif-negatif atau model pembelajaran mesin yang dilatih sebelumnya. Subjektivitas mengukur sejauh mana teks mengungkapkan pendapat pribadi atau opini; teks yang sangat subjektif lebih mengandung opini, sementara teks yang objektif lebih netral dan tidak mengandung opini pribadi.

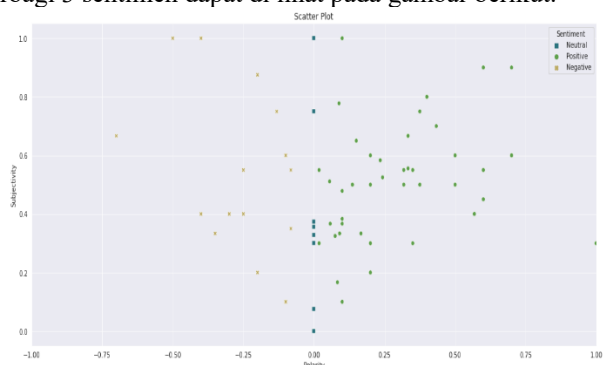
Dalam analisis sentimen ini, penting untuk memahami bahwa polaritas dan subjektivitas bekerja bersama untuk menentukan sentimen keseluruhan suatu teks. Sebuah ulasan bisa saja mengandung kata-kata positif, namun jika disampaikan dengan opini yang kuat, dapat berkontribusi pada klasifikasi sebagai sentimen positif. Sebaliknya, ulasan yang lebih netral atau objektif, meski berisi kata-kata positif, akan diklasifikasikan sebagai netral. Dengan menggunakan kedua faktor ini, model dapat dengan lebih akurat mengklasifikasikan sentimen sebagai positif, negatif, atau netral berdasarkan pola yang ditemukan dalam teks ulasan.:

Dengan memperhatikan kedua faktor tersebut, yaitu polaritas dan subjektivitas, model berhasil mengklasifikasikan setiap dokumen ke dalam kategori sentimen yang sesuai. Berikut adalah hasil klasifikasi sentimen dari masing-masing dokumen yang mencakup sentimen positif, negatif, dan netral dalam tabel berikut.

TABEL VIII
ACUAN UKURAN TEKS

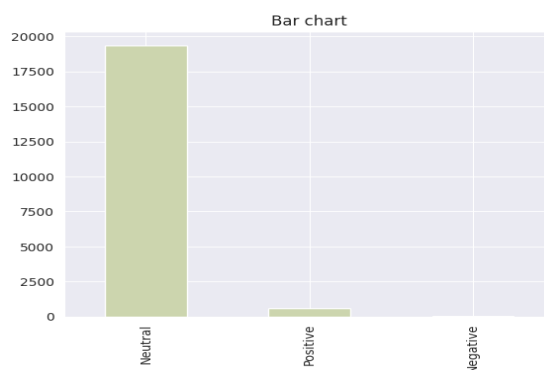
Dokumen Sentiment	Kata	Nilai	Polarity	Subjectivity
Positif (6892)	wow	4.00	0.10	1.00
	aplikasi	0.00		
	keren	0.00		
Negatif (18338)	pusing	-0.10	-0.10	0.20
	mental	0.00		
	ya	0.00		
Netral (15225)	pingin	0.00	0.00	0.00
	bisnis	0.00		
	online	0.00		
	tau	0.00		
	buka	0.00		
	lapak	0.00		

Hasil dari melalui metode polarity dan subjectivity yang terbagi 3 sentimen dapat dilihat pada gambar berikut.



Gambar 2. Persebaran Subjectivity Dan Polarity.

Selanjutnya dari table VII divisualisasikan dengan asngka setiap sentiment sebagai berikut.



Gambar 3. Visualisasi Hasil angka Pada Setiap Sentimen.

Berdasarkan dari data pada gambar 2 dan gambar 3 diatas, sentiment netral mendapatkan 1031 data hasil kemudian pada

sentiment positive mendapatkan 219 data dan sentiment negatif mendapatkan hasil sangat kecil yaitu 50 data.

C. Hasil TF-IDF

Pada tahap TF-IDF, pembobotan dilakukan untuk mengukur seberapa sering kata muncul dalam dokumen (TF) dan seberapa penting kata tersebut di seluruh koleksi dokumen (IDF). Nilai TF-IDF dihitung dengan mengalikan TF dan IDF untuk setiap kata dalam dokumen. Semakin tinggi nilai TF-IDF, semakin relevan kata tersebut dalam menggambarkan sentimen dalam dokumen. Setelah perhitungan bobot dilakukan, kata-kata yang memiliki TF-IDF tertinggi dianggap paling relevan terhadap sentimen yang diekspresikan dalam dokumen.

```
dokumen = term_fit.transform(Ulasan) #hasil perhitungan tf idf dalam 1 doc
tfidf_transformer = TfidfTransformer().fit(dokumen)
print (tfidf_transformer.idf_)

tfidf=tfidf_transformer.transform(term_frequency)
print (tfidf) #hasil manual dengan sistem pyhton

[10.21039037 10.21039037 10.21039037 ... 10.21039037 10.21039037
 10.21039037]
(0, 3089)      0.6699190942450138
(0, 1980)      0.7424341096460616
```

Sebagai contoh, pada dokumen sentimen positif (6892), kata "komplain" memiliki nilai TF-IDF tertinggi (0.142), yang menunjukkan kata ini sangat relevan untuk menggambarkan sentimen positif. Pada dokumen sentimen negatif (18338), kata "kalo" memiliki nilai TF-IDF tertinggi (0.121), mengindikasikan relevansi dalam sentimen negatif. Sementara itu, pada dokumen sentimen netral (15225), kata-kata seperti "bayar" dan "biaya" memiliki TF-IDF yang rendah, yang menandakan bahwa dokumen ini bersifat lebih objektif dan tidak mengungkapkan sentimen kuat.

D. Implementasi Multinomial Naïve Bayes

Sebelum implementasi Multinomial Naïve Bayes, data dibagi menjadi dua bagian: data latih dan data uji. Data latih digunakan untuk melatih model, sedangkan data uji digunakan untuk menguji akurasi model yang telah dilatih. Proses pembagian data ini penting untuk menghindari overfitting, di mana model mungkin terlalu menyesuaikan diri dengan data latih dan tidak dapat memprediksi dengan baik pada data baru.

```
from sklearn.model_selection import train_test_split

X_train, X_test, y_train, y_test = train_test_split(data_label['Ulasan_siap'],
                                                    data_label['Sentiment'],
                                                    test_size=0.1,
                                                    stratify=data_label['Sentiment'],
                                                    random_state=30)
```

Setelah pembagian, model Multinomial Naïve Bayes diterapkan pada data latih untuk membangun model klasifikasi berdasarkan kata-kata dalam ulasan. Kemudian, model diuji menggunakan data uji untuk mengevaluasi performanya dalam mengklasifikasikan sentimen (positif, negatif, atau netral) berdasarkan TF-IDF yang telah dihitung sebelumnya.

```
# Apply undersampling to balance the classes
rus = RandomUnderSampler(sampling_strategy='auto', random_state=42)
X_train_resampled, y_train_resampled = rus.fit_resample(X_train, y_train)

# Train the Gaussian Naive Bayes classifier on the resampled training set
gnb = GaussianNB()
gnb.fit(X_train_resampled.toarray(), y_train_resampled)

# Evaluate the classifier on the testing set
y_pred = gnb.predict(X_test.toarray())
accuracy = accuracy_score(y_test, y_pred)
print('Accuracy:', accuracy)

# Print classification report and confusion matrix
cm = confusion_matrix(y_test, y_pred)
print(classification_report(y_test, y_pred))

# Use the classifier to predict the sentiment class for all reviews in the dataset
data_label['predicted_sentiment_class'] = gnb.predict(X.toarray())

# Save the resulting dataframe to a new CSV file
data_label.to_csv('predicted_sentiment1.csv', index=False)
```

Selanjutnya, dengan syntax from sklearn.metrics import accuracy_score accuracy = accuracy_score(y_test, y_pred) untuk melihat dan menyimpan akurasi skor pada dataset.

```
Accuracy: 0.8051666666666667
Confusion Matrix:
[[ 5  1  1]
 [425 4694 703]
 [ 9  30 132]]
Classification Report:
              precision    recall  f1-score   support

 Negative         0.01         0.71         0.02         7
  Neutral         0.99         0.81         0.89        5822
   Positive         0.16         0.77         0.26         171

 accuracy                   0.81        6000
 macro avg              0.39         0.76         0.39        6000
 weighted avg           0.97         0.81         0.87        6000
```

Gambar 4 Evaluasi Model

Evaluasi model dilakukan menggunakan metrik accuracy score, precision score, recall score, dan f1-score. Hasil evaluasi menunjukkan bahwa untuk sentimen negatif, model menghasilkan precision sebesar 0.1, recall 0.71, dan F1-score 0.02, yang menunjukkan kesulitan model dalam mengklasifikasikan sentimen negatif dengan benar. Untuk sentimen netral, model menunjukkan precision 0.99, recall 0.81, dan F1-score 0.89, yang menunjukkan bahwa model sangat baik dalam mengklasifikasikan ulasan yang tidak mengandung sentimen kuat. Sedangkan untuk sentimen positif, model menghasilkan precision 0.09, recall 0.30, dan F1-score 0.26, yang menunjukkan bahwa model kesulitan dalam mengenali sentimen positif dengan akurasi tinggi.

Berdasarkan confusion matrix yang diperoleh, model menunjukkan akurasi sebesar 81%, dengan distribusi hasil sebagai berikut: True Positive (TP) untuk sentimen negatif adalah 1, netral 136, dan positif 20. False Positive (FP) untuk

sentimen negatif adalah 7, netral 17, dan positif 12. False Negative (FN) untuk sentimen negatif adalah 4, netral 62, dan positif 1. Hasil ini menunjukkan bahwa model lebih efektif

dalam mengklasifikasikan sentimen netral, namun kesulitan dalam mengklasifikasikan sentimen positif dan negatif dengan akurasi yang lebih rendah.



Gambar 5. Visualisasi Sentimen (a. Visualisasi Sentimen Positif, b. Visualisasi Sentimen Negatif)

Berdasarkan Gambar 5, Visualisasi Sentimen Positif menunjukkan kata-kata seperti "mantap", "sukses", "banget", "ok barang", "wow", "online", "jual", "mudah", "free", "ongkir", "akun", "terima kasih", "percaya", "layan", "bukti", "baik", "lindung", "biaya", dan "seller". Meskipun kata-kata seperti "free" dan "ongkir" memiliki bobot rendah, kata-kata positif lainnya seperti "mudah", "aman", dan "bagus" memberi pengaruh kuat terhadap sentimen positif. Visualisasi ini menunjukkan bahwa komentar dengan sentimen positif tetap memberikan saran konstruktif untuk perbaikan, meskipun ada keluhan minor.

Visualisasi Sentimen Negatif, di sisi lain, menunjukkan kata-kata seperti "beli", "tipu", "minimal", "jual", "voucher", "mending", "admin", "system", "ongkir", "sampah", "susah", "enggak", "untung", dan "update". Komentar negatif lebih banyak berfokus pada masalah sistem pembelian dan pengiriman, mengindikasikan ketidakpuasan terhadap proses transaksi dan pelayanan yang memerlukan perhatian lebih untuk target marketing dan perbaikan di masa depan.



Gambar 6. Visualisasi Sentimen Netral

Gambar 6 menunjukkan bahwa sentimen netral mencakup kata-kata seperti "manfaat", "bantu", "terimakasih", "belanja parah", "sukses", "layan", "puas", "gratis", "ongkir", "gampang", "bagus", "percaya", "amanah", "mending", "hilang", "keren", "respon", dan "cepat". Komentar pada sentimen netral bervariasi, dengan beberapa kata yang menunjukkan elemen positif seperti "mudah", "aman", dan "bagus", sementara kata-kata lain seperti "kecewa" dan "susah" menunjukkan nuansa negatif. Hal ini menunjukkan bahwa sentimen netral sering kali menggabungkan elemen dari kedua sentimen positif dan negatif, mencerminkan keseimbangan atau ketidakpastian dalam pengalaman pengguna.

	Nama pengguna	Ulasan_clean	Sentiment	predicted_sentiment_class
0	Nizar Rayyan	paylater batal otomatis kadang nya ganti jatuh...	Neutral	Neutral
1	Fandy Komo	kesini buka lapak bagus layan komplan susah su...	Neutral	Neutral
2	dedy kelana (CV PRIBUMI BANGKIT)	pasti hp wifi lemmoootttt dasar bukalapak kar...	Positive	Positive
3	hidayat mudaris	tolong rumit verifikasi akun saldo bukalapak p...	Negative	Negative
4	ari sandy	aplikasi wadah tipu gampang banget tipu tim sy...	Neutral	Neutral
...
1295	Pengguna Google	aplikasi berat bos saran banyak fitur baik per...	Positive	Positive
1296	Pengguna Google	aplikasi bukalapak kecewa order customer kurir...	Neutral	Neutral
1297	Roofie Yk	paylaterpadahal tingkat belanja non cod mohon ...	Neutral	Neutral
1298	Anto Suhendra	tingkat aman nya ya pakai bukalapak betahun pi...	Neutral	Neutral
1299	Erwin Rusdiyanto	haus az tuh paylaternya paylater pakai	Neutral	Neutral

1300 rows x 4 columns

Gambar 7 Prediksi Hasil Multinomial Naïve Bayes

Data predict pada gambar 5 di hasilkan berdasarkan dari data latih dan data uji melakukan prediksi mendapatkan hasil Neutral 868 Positive 326 dan Negative 106 cukup memiliki perbedaan dengan data di miliki pada saat identifikasi sentimen dengan polarity dan subjectivity dengan hasil Neutral 1056 Positive 198 dan Negative 46 yaitu neutral mengalami pengurangan dan positif mengalami kenaikan dan juga negative mengalami kenaikan.

Selanjutnya, hasil klasifikasi sentimen menggunakan metode Multinomial Naïve Bayes menunjukkan performa yang cukup baik dalam mengklasifikasikan ulasan pengguna Bukalapak. Sentimen netral mendapat akurasi yang tinggi, dengan F1-score yang mencapai 0.89, yang mengindikasikan bahwa model dapat dengan baik membedakan ulasan yang tidak memiliki perasaan kuat, meskipun ada campuran kata positif dan negatif di dalamnya. Hal ini menunjukkan bahwa banyak ulasan dari pengguna yang tidak mengungkapkan sentimen yang ekstrem, yang dapat mengarah pada kebutuhan Bukalapak untuk fokus pada peningkatan pengalaman pelanggan secara keseluruhan, bukan hanya pada ulasan yang sangat positif atau negatif.

Namun, model ini menunjukkan kelemahan dalam mengklasifikasikan sentimen positif dan negatif dengan akurasi yang lebih rendah, terutama pada sentimen positif yang hanya memiliki F1-score sebesar 0.26. Hal ini mungkin disebabkan oleh ketidakseimbangan data, di mana ulasan negatif lebih dominan dalam dataset. Ulasan negatif sering kali lebih jelas dan ekspresif, sementara ulasan positif cenderung lebih netral dan tidak se-ekspresif ulasan negatif, sehingga model kesulitan dalam mengidentifikasi sentimen positif secara akurat.

Hasil ini memberikan gambaran yang jelas tentang bagaimana Bukalapak dapat memanfaatkan analisis sentimen untuk meningkatkan pengelolaan produk dan layanan mereka. Dengan mengenali ulasan yang netral, perusahaan dapat lebih fokus untuk memperbaiki aspek layanan yang dapat meningkatkan kepuasan pelanggan tanpa mengandalkan umpan balik ekstrem (positif atau negatif). Selain itu, analisis lebih lanjut terhadap sentimen negatif dapat membantu Bukalapak untuk menangani masalah yang paling mendesak

bagi pelanggan, seperti kendala dalam pengiriman, masalah sistem, dan pelayanan pelanggan yang buruk. Di sisi lain, sentimen positif yang kurang terdeteksi dengan baik oleh model ini memberikan ruang bagi Bukalapak untuk menyempurnakan teknik pengenalan sentimen positif dengan memperhatikan elemen-elemen positif yang lebih halus dalam ulasan.

IV. KESIMPULAN

Penelitian ini berhasil menerapkan metode Multinomial Naïve Bayes untuk analisis sentimen terhadap 19.999 ulasan Bukalapak selama tahun 2023. Hasil penelitian menunjukkan bahwa model ini mencapai akurasi sebesar 80,52%, dengan sentimen netral diklasifikasikan dengan sangat baik (precision 0.99, recall 0.81, F1-score 0.89), sementara sentimen positif dan negatif membutuhkan perbaikan lebih lanjut. Kontribusi utama penelitian ini adalah penerapan penanganan ketidakseimbangan data menggunakan SMOTE dan penerapan k-fold cross-validation yang meningkatkan keandalan model dalam mengklasifikasikan sentimen pada data yang besar dan tidak seimbang. Perbaikan metode yang dilakukan dalam penelitian ini, termasuk pengolahan data teks dengan teknik normalisasi, tokenisasi, stemming, serta penggunaan TF-IDF, menunjukkan peningkatan yang signifikan dalam memahami dan mengklasifikasikan sentimen pengguna. Penelitian ini juga menyarankan untuk mengatasi tantangan linguistik dalam Bahasa Indonesia, seperti slang, imbuhan, dan negasi, yang penting untuk meningkatkan akurasi dalam analisis sentimen Bahasa Indonesia.

Hasil analisis sentimen ini dapat memberikan manfaat langsung bagi platform e-commerce seperti Bukalapak. Dengan menganalisis ulasan pengguna secara otomatis, Bukalapak dapat lebih mudah mengidentifikasi area yang membutuhkan perbaikan, seperti layanan pelanggan, fitur aplikasi, dan pengelolaan produk. Penelitian ini juga menunjukkan pentingnya mengatasi sentimen negatif untuk meningkatkan kepuasan pelanggan dengan memperbaiki aspek-aspek seperti pengembalian barang dan voucher belanja. Secara keseluruhan, penelitian ini memberikan kontribusi penting dalam meningkatkan pengelolaan dan layanan di platform e-commerce melalui analisis sentimen yang lebih akurat dan efisien.

DAFTAR PUSTAKA

- [1] N. Adiasa, "Pengaruh Pemahaman Peraturan Pajak terhadap Kepatuhan Wajib Pajak dengan Moderating Preferensi Risiko," *Accounting Analysis Journal*, vol. 2, no. 3, pp. 345–352, 2013.
- [2] S. Andrews and L. Hirsch, "A Tool for Creating and Visualising Formal Concept Trees," *CEUR Workshop Proceedings*, vol. 1637, pp. 1–9, 2016.
- [3] A. Agustinah, "Word Cloud of Corruption Eradication Commission," pp. 4–5, 2015.
- [4] A. Aswin and A. Wahidun, "Analisis Atribut Produk Samsung dan Asus Menggunakan Metode Multidimensional Scaling (MDS) di Bandar Lampung," *Jurnal Bisnis Darmajaya*, vol. 2, no. 2, pp. 62–74, 2016.

- [5] R. Feldman and J. Sanger, *The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data*, Cambridge University Press, 2007.
- [6] F. Fitria and I. Dwijananda, "Analisis Pengaruh Electronic Word of Mouth terhadap Proses Keputusan Pembelian (Studi pada Go-Jek)," *Eproceeding of Management*, vol. 3, pp. 1–19, 2016.
- [7] M. Rasyadi, "Analisis Sentimen pada Twitter Menggunakan Metode Naïve Bayes (Studi Kasus Pemilihan Gubernur DKI Jakarta 2017)," pp. 1–17, 2017.
- [8] C. Sagita, "Pengaruh Electronic Word of Mouth, Brand Ambassador, dan Persepsi Nilai terhadap Keputusan Pembelian pada Tokopedia.com," IIB Darmajaya, 2020. [Online]. Available: <http://repo.darmajaya.ac.id/id/eprint/2673>
- [9] F. Gorunescu, *Data Mining: Concepts, Models and Techniques*, Springer Science & Business Media, 2011.
- [10] C. Kaur and A. Sharma, "Twitter Sentiment Analysis on Coronavirus Using TextBlob," *EasyChair Preprint* 2974, pp. 1–10, 2020.
- [11] S. Kim, K. Han, H. Rim, and S. Myaeng, "Some Effective Techniques for Naive Bayes Text Classification," *IEEE Transactions on Knowledge and Data Engineering*, vol. 24, pp. 1457–1466, 2006.
- [12] N. Komang et al., "Seleksi Fitur Bobot Kata dengan Metode TF-IDF untuk Ringkasan Bahasa Indonesia," *Merpati*, vol. 6, no. 2, 2018.
- [13] A. Kurniawan, "Analisis Kondisi Lingkungan Fisik Rumah dengan Kejadian ISPA pada Balita di Wilayah Puskesmas Purwokerto Selatan Kecamatan Purwokerto Selatan Kabupaten Banyumas Tahun 2013," Universitas Harapan Bangsa, 2013. Available: <http://eprints.uhb.ac.id/id/eprint/2103>
- [14] C. K. Laudon and P. J. Laudon, *Essentials of Management Information Systems*, Pearson Education, Inc., 2013.
- [15] M. Mohri, A. Rostamizadeh, and A. Talwalkar, *Foundations of Machine Learning*, 2012.
- [16] A. Mustika and M. Affandes, "Penerapan Metode Support Vector Machine dalam Klasifikasi Sentimen Tweet Public Figure," *Sentra*, pp. 978–979, 2015.
- [17] D. Normawati and S. A. Prayogi, "Implementasi Naïve Bayes Classifier dan Confusion Matrix pada Analisis Sentimen Berbasis Teks pada Twitter," *Jurnal Sains Komputer & Informatika (J-Sakti)*, vol. 5, no. 2, pp. 697–711, 2021.
- [18] F. Nurhuda, S. W. Sihwi, and A. Doewes, "Analisis Sentimen Masyarakat terhadap Calon Presiden Indonesia 2014 Berdasarkan Opini dari Twitter Menggunakan Metode Naïve Bayes Classifier," *ITSmart: Jurnal Teknologi dan Informasi*, vol. 2, no. 2, pp. 35–42, 2013.
- [19] L. Perkovic, *Introduction to Computing Using Python*, pp. 510, 2012. Available: https://dspace.uii.ac.id/bitstream/handle/123456789/7762/14611242_syarifah_rosita_dewi_statistika.pdf?Sequence=1
- [20] P. Kotler and K. L. Keller, *A Framework for Marketing Management (Sixth Edition-Global Edition)*, 2016.
- [21] J. Pustejovsky and A. Stubbs, *Natural Language Annotation for Machine Learning: A Guide to Corpus-Building for Applications*, 2012. Available: <https://books.google.co.id/books?id=nm4v1k0yc>
- [22] M. Rasyadi, "Analisis Sentimen pada Twitter Menggunakan Metode Naïve Bayes (Studi Kasus Pemilihan Gubernur DKI Jakarta 2017)," pp. 1–17, 2017.
- [23] C. Sagita, "Pengaruh Electronic Word of Mouth, Brand Ambassador, dan Persepsi Nilai terhadap Keputusan Pembelian pada Tokopedia.com," IIB Darmajaya, 2020. Available: <http://repo.darmajaya.ac.id/id/eprint/2673>
- [24] U. Sumarwan, U. Simanjuntak, and L. N. Yuliati, "Meta-Analysis Study: Reading Behavior of Food Products Label," *Journal of Consumer Sciences*, vol. 2, no. 2, pp. 26, 2017.