



## Perbandingan Algoritma K-Nearest Neighbor dan Naïve Bayes pada Aplikasi Shopee

A. Oktian Permana<sup>\*1</sup>, Sudin Saepudin<sup>2</sup>

Email: <sup>1</sup>a.oktian\_si19@nusaputra.ac.id, <sup>2</sup> sudin.saepudin@nusaputra.ac.id

<sup>1,2</sup>Program Studi Sistem Informasi, Universitas Nusa Putra

Diterima: 4 Januari 2023 | Direvisi: 28 April 2023 | Disetujui: 28 Mei 2023  
©2020 Program Studi Teknik Informatika Fakultas Ilmu Komputer,  
Universitas Muhammadiyah Riau, Indonesia

### Abstrak

E-commerce adalah aplikasi untuk melakukan semua aktivitas jual beli secara online ada juga beberapa masalah dalam transaksi penjualan online, terutama bagi pembeli, yang mungkin menemui banyak kendala. Ulasan dari pengguna sering digunakan sebagai alat yang efektif dan efisien untuk mencari informasi tentang suatu produk atau layanan. Pengguna internet sering melihat ulasan suatu produk sebelum menggunakannya, karena ulasan dari pengguna lain dapat memberikan informasi terkini. Pada penelitian ini penulis menggunakan dua (2) jenis metode algoritma yakni K-Nearest Neighbor dan Naïve Bayes Classifier. Penulis menggunakan dua (2) metode untuk mengetahui tingkat akurasi terbaik yang di hasilkan. Penulis menggunakan algoritma K-Nearest Neighbor karena pada penelitian sebelumnya terdapat hasil yang bagus. Juga menggunakan algoritma Naïve Bayes Classifier dikarenakan metode ini merupakan metode populer dan lebih sering digunakan. Algoritma K-Nearest Neighbor pada aplikasi shopee menunjukkan bahwa tingkat akurasi 55.00% dimana jumlah data untuk sentimen negatif sebanyak 100 ulasan, dan sentimen positif sebanyak 100 ulasan dari data testing yang digunakan sebanyak 200 data atau 20% dalam metode K-Nearest Neighbor. Lalu tingkat akurasi mencapai 80,00% dimana jumlah data untuk sentimen negatif sebanyak 99 ulasan dan sentimen positif sebanyak 101 ulasan dari data testing yang digunakan sebanyak 200 atau 20% pada metode Naïve Bayes Classifier. Untuk penelitian yang telah dilakukan peneliti, bisa dilihat bahwa metode Naïve Bayes Classifier merupakan metode dengan tingkat akurasi yang lebih tinggi dibandingkan K-Nearest Neighbor dengan tingkat akurasi sebesar 80,00%

**Kata kunci:** Analisis Sentimen, Belanja Online, K-Nearest Neighbor, Naïve Bayes Classifier

### *Comparison of the K-Nearest Neighbor and Naïve Bayes Algorithms on the Sentiment Analysis Shopee Application*

#### *Abstract*

*E-commerce is an application for carrying out all buying and selling activities online. There are also some problems in online sales transactions, especially for buyers, who may encounter many obstacles. Reviews from users are often used as an effective and efficient tool for finding information about a product or service. Internet users often see reviews of a product before using it, because reviews from other users can provide up-to-date information. In this study the authors used two (2) types of algorithm methods namely K-Nearest Neighbor and Naïve Bayes Classifier. The author uses two (2) methods to determine the best level of accuracy produced. The author uses the K-Nearest Neighbor algorithm because in previous studies there were good results. Also uses the Naïve Bayes Classifier algorithm because this method is a popular method and is used more often. The K-Nearest Neighbor algorithm in the shopee application shows that the accuracy rate is 55.00% where the amount of data for negative sentiment is 100 reviews, and positive sentiment is 100 reviews from data testing used as much as 200 data or 20% in the K-Nearest Neighbor method. Then the accuracy level reaches 80.00% where the amount of data for negative sentiment is 99 reviews and positive sentiment is 101 reviews from the testing data used as much as 200 or 20% in the Naïve Bayes Classifier method. For research that has been conducted by researchers, it can be seen that the Naïve Bayes Classifier method is a method with a higher level of accuracy than the K-Nearest Neighbor with an accuracy rate of 80.00%*

**Keywords:** Sentiment Analyze, Online Shopping, K-Nearst Neighbor, Naive Bayes Classifier

---

## 1. PENDAHULUAN

Era *modernisasi* menjadikan manusia saling berlomba menciptakan sebuah teknologi untuk mempermudah setiap aktivitas kesehariannya dengan semakin majunya teknologi masyarakat mulai banyak yang beralih untuk melakukan pembelian secara digital yang banyak tersedia di era sekarang ini. *E-commerce* adalah aplikasi untuk melakukan semua aktivitas jual beli secara *online* [1].

Shopee ini adalah jenis *e-commerce consumer-to-consumer (C2C)*. CEO Shopee adalah Chris Feng. Shopee co.id diluncurkan di Indonesia pada Juni 2015[2]. Ditambah dengan metode pembayaran yang telah banyak didukung oleh dompet digital yang sudah marak di masyarakat. Dompet digital (*E-wallet*) saat ini berkembang begitu pesat, Apalagi banyak penawaran menarik sehingga masyarakat sudah mulai beralih ke *e-wallet* untuk melakukan transaksi keuangan. Dompet digital (*e-wallet*) dapat digunakan untuk membayar berbagai transaksi yang tersedia. Pengguna dapat melakukan transaksi keuangan tanpa uang tunai *online* dan *offline*. Penerapan sistem non tunai di Indonesia merupakan pengaruh langsung dari perkembangan *financial technology (Fintech)*. Penggunaan *e-wallet* dinilai berdampak positif karena lebih efisien dan berdampak pada berkurangnya perputaran uang tunai. Di Indonesia sudah banyak layanan *e-wallet* seperti OVO, DANA, Go-Pay dan masih banyak lagi.

Menurut Veitzhal Rivai (2012: 237), ada juga beberapa masalah dalam transaksi penjualan online, terutama bagi pembeli, yang mungkin menemui banyak kendala, seperti (a) kualitas barang yang dijual; barang untuk dibeli. Pembeli melihat gambar tampilan produk yang dijual dan tidak dapat memastikan secara spesifik apakah produk tersebut memenuhi spesifikasi atau kriteria pembelian. (b) perkiraan waktu penyerahan barang menarik perhatian pembeli terhadap produk atau barang yang akan dibeli; (c) pada saat barang diserahkan kepada pembeli, barang yang diterima tidak sesuai dengan barang yang ditawarkan di website Toko Online, terlepas dari spesifikasi, jenis atau sifat barang yang dijual; (d) Apabila pembeli mengajukan hak klaim (pengembalian dana) ke website toko online, maka tidak semua penjual toko online bersedia menanggapi barang yang diklaim [3].

Ulasan dari pengguna sering digunakan sebagai alat yang efektif dan efisien untuk mencari informasi tentang suatu produk atau layanan. Pengguna internet sering melihat ulasan suatu produk sebelum menggunakannya, karena ulasan dari pengguna lain dapat memberikan informasi terkini tentang produk tersebut. Namun tidak mudah untuk memantau dan mengatur *opini* publik. Terlalu banyak *opini* yang *diposting* di media sosial untuk diproses secara manual. Oleh karena itu, peneliti mengkaji sejauh mana analisis ulasan pengguna terhadap aplikasi Shopee berdasarkan jumlah bintang yang diberikan setelah menggunakan aplikasi tersebut.

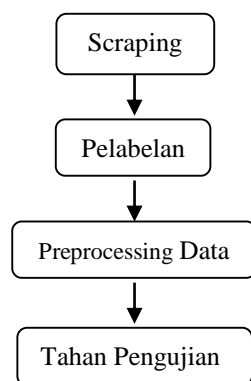
Analisis sentimen dalam konteks ini dapat digunakan untuk mempelajari pandangan, perilaku, dan perasaan atau emosi seseorang saat mengungkapkan pendapat. Besarnya dampak serta manfaat analisis sentimen telah menyebabkan pertumbuhan pesat dalam penelitian dan aplikasi berdasarkan analisis sentimen.

Dari penggunaan metode Naive Bayes yang dilakukan oleh [4] menyimpulkan bahwa itu dapat digunakan untuk memahami persepsi *internasional* tentang *tagar* yang populer di media sosial. Juga pada penelitian lainnya yang dilakukan oleh [5] dapat ditarik kesimpulan Algoritma K-Nearst Neighbor mempunyai kinerja yang baik dalam klasifikasi data. Pada penelitian yang lain yang dilakukan oleh [6] mendapatkan hasil yang memuaskan untuk algoritma *naive bayes*.

Pada penelitian ini akan dibahas langkah-langkah dalam proses melakukan analisis sentimen terhadap review dari sebuah aplikasi shopee Dari tahap *preprocessing* hingga tahap analisis sentimen menggunakan K-Nearst Neighbor dan Naive Bayes Classifier.

## 2. METODE PENELITIAN

Metode meliputi analisis masalah dan desain untuk pemecahan masalah. Analisis menggambarkan metodologi penelitian sebagai suatu proses yang harus ada agar penelitian dapat diorganisasikan untuk memperoleh hasil yang sesuai dengan tujuan penelitian. Tahapan pendekatan penelitian ditunjukkan pada Gambar 1.



Gambar 1 Tahapan Penelitian

### 2.1. Scraping Data

*Web scraping* merupakan metode untuk memperoleh data dari situs web. Pada penelitian ini *scraping* dilakukan pada google playstore dengan *tools colab google* yang merupakan produk dari google menggunakan algoritma *python*. Data yang dikumpulkan adalah ulasan aplikasi shopee yang tersedia di playstore, kemudian disimpan dalam *CSV (Comma Separated Values)* untuk digunakan nanti pada *preprocessing data*, dengan data hingga 1000 data ulasan.

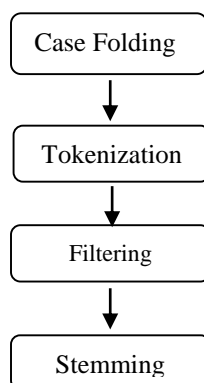
### 2.2. Pelabelan Otomatis

Tahap berikutnya setelah data tersimpan pada bentuk *CSV (Comma Separated Values)* selanjutnya adalah pemberian label yang dilakukan secara otomatis dengan *script* pada algoritma *python*.

### 2.3. Preprocessing Data

Kemudian dilanjutkan dengan *preprocessing data*. *Preprocessing* Preprocessing adalah proses mempersiapkan data sebelum pemodelan. Preprocessing adalah salah satu teknik dalam data mining untuk mengubah data mentah menjadi format yang lebih mudah dipahami [7]. Proses ini merupakan tujuan agar data bisa diklasifikasi untuk mempermudah proses analisis menggunakan algoritma K-Nearest Neighbor dan Naive Bayes Classifier.

pada tahap ini data akan melalui beberapa *preprocessing* sebelum memasuki tahap untuk proses *validasi* yang diawali dengan *case folding, tokenize, filtering* dan *stemming*. Berikut ini adalah proses tahapan *preprocessing* yang bisa terlihat pada gambar 2.



Gambar 2 Tahapan Preprocessing

- 1) *Case Folding* : *Case folding* adalah proses mengubah setiap kata dalam dataset menjadi huruf kecil menggunakan fungsi (*lowercase*) [8].
- 2) *Tokenization* : *Tokenization* dan merupakan proses penguraian teks kalimat menjadi kata-kata [8].
- 3) *Filtering* : *filtering* merupakan proses menghilangkan data yang tidak lengkap, data yang error [7].
- 4) *Stemming* : *Stemming* merupakan proses pencarian kata dasar, dan juga menghilangkan semua imbuhan yang terdapat pada suatu kata [7]. merupakan proses pencarian kata pada suatu dokumen untuk mengetahui berapa jumlah kata , yang kemudian diberi bobot menggunakan TF-IDF.

## 2.4. Metode Klasifikasi

Untuk penelitian ini penulis memakai dua (2) jenis metode algoritma yakni K-Nearst Neighbor dan Naive Bayes Classifier. Penulis menggunakan dua (2) metode untuk mengetahui tingkat akurasi terbaik yang di hasilkan. Penulis menggunakan algoritma K-Nearst Neighbor karena pada penelitian sebelumnya terdapat hasil yang bagus. Juga penulis menggunakan algoritma Naive Bayes Classifier dikarenakan metode ini merupakan metode yang populer dan lebih sering digunakan.

K-Nearst Neighbor adalah salah satu algoritma paling sederhana untuk memecahkan masalah klasifikasi. Algoritma ini sering digunakan dalam klasifikasi teks dan data. Dalam pendekatan ini, objek diklasifikasikan berdasarkan data yang paling dekat dengan objek tersebut [9].

Algoritma dirancang untuk mengklasifikasikan objek berdasarkan atribut dan sampel data pelatihan. Seperti K-means, metode ini sangat sederhana, tidak membuat asumsi tentang distribusi data, mudah diimplementasikan, dan sering digunakan dalam kasus-kasus praktis. Langkah-langkah algoritma K-NN adalah sebagai berikut:

1. Tentukan nilai K
2. Hitung jarak antara data baru dengan semua data training
3. Urutkan jarak terpendek
4. Periksa kelas yang berdekatan K
5. Kelas data baru = kelas mayoritas tetangga terdekatnya

Perhitungan K-NN adalah menjumlahkan semua nilai kesamaan yang terdapat pada suatu kategori dan membandingkan mana yang lebih besar. Persamaan K-NN adalah sebagai berikut:

$$\cos Sim(A, B) = \frac{\sum_{i=1}^n A_i B_i}{\sqrt{(\sum_{i=1}^n A_i)^2} \sqrt{(\sum_{i=1}^n B_i)^2}}$$

A = Data uji, B = Data latih, Ai dan B = bobot nilai yang diberikan untuk setiap *term* yang ada[10].

Algoritma naive bayes classifier adalah sebuah algoritma teknik klasifikasi yang memanfaatkan metode probabilistik dan statistik. Algoritma ini merupakan salah satu metode klasifikasi yang populer, menurut IEEE International Conference on Data Mining (ICDM'06) di Hongkong dan termasuk dalam sepuluh besar algoritma dalam bidang data mining[11]. Naive Bayes merupakan cabang matematika yang dikenal sebagai teori probabilitas untuk mencari peluang terbesar dari kemungkinan klasifikasi, dengan melihat frekuensi masing-masing klasifikasi dalam data pelatihan [12]. Multinomial Naive Bayes menilai apakah data input memiliki kelas tertentu dengan menghitung probabilitas posterior (prior). Dengan persamaan sebagai berikut :

$$P(C_i|d) = P(C_i) \prod_{k=1}^{n_d} P(t_k | C_i)$$

P(Ci|d) adalah probabilitas posterior dari Ci diberikan dokumen d, Ci adalah class variable ke i,  $i = \{1, \dots, n\}$ , P(Ci) adalah probabilitas posterior kelas ke-I, P(tk|Ci) adalah probabilitas dari t ke-k, diberikan kelas C ke i. P(tk|Ci) disebut sebagai likelihood. dan nd adalah jumlah fitur yang terdapat dalam dokumen d[13].

Dalam penelitian ini, peneliti menggunakan fitur NLTK untuk mengklasifikasikan text mining. NLTK merupakan *tools* paling sering untuk digunakan pada pemrosesan data di ilmu Natural Language Processing (NLP) dengan menerapkan penggunaan bahasa pemrograman *Python*. Natural Language Toolkit juga mendukung untuk pemrosesan bahasa alamiah seperti *classification, tokenization, stemming, tagging, parsing* dan *filtering* [4].

## 3. HASIL DAN PEMBAHASAN

Hasil dan pembahasan untuk analisis sentimen pada aplikasi shopee menggunakan pendekatan naive bayes classifier yang diolah oleh peneliti menggunakan *Google Collab*, contoh data *scraped*, berlabel, *preprocessing* dan nilai akurasi untuk masing-masing pendekatan disajikan pada sub bab selanjutnya.

### 3.1. Scraping data

Hasil hari scraping data yang telah dilakukan dengan bobot data pencarian sebanyak 1000 data pada masing-masng aplikasi dapat dilihat pada gambar dibawah ini dengan menggunakan *script*.

Scraping

```
from google_play_scraper import Sort, reviews

result, continuation_token = reviews(
    'id.dana',
    lang='id',
    country='id',
    sort=Sort.NEWEST,
    count=1000,
    filter_score_with=None
)
```

	userName	score	at	content
0	revina Safitri	2	2022-11-25 06:48:16	Sumpah lemot banget please, sering nge bug git...
1	Auliarubi Kusmawan	5	2022-11-25 06:47:52	Bagusssssssssssssssssssssssssssssssssss
2	John Bay	5	2022-11-25 06:47:47	Joshh
3	Davit Irawan	5	2022-11-25 06:46:22	Mantap 🤩
4	rina wahyu	5	2022-11-25 06:46:08	Shopee kok sering macet ya, jd harus sering tu...
...	...	...	...	...
995	Nabila Zulaikha	1	2022-11-24 12:08:56	APA SIH MAKIN HARI SHOPEE UDAH ENGGAK KAYA DUK...
996	Mohamad Silva Sulhana	3	2022-11-24 12:08:55	ongkirnya lbh mahal drpd barang yg diberi ya, ...
997	Syafira Qiandra	1	2022-11-24 12:08:05	Duh kenapa ada biaya macam macam sekarang apal...
998	willy kris Hadiyanto	5	2022-11-24 12:07:52	Mudah dan keren abis
999	Suyatmi Sby	5	2022-11-24 12:07:22	Memuaskan

1000 rows x 4 columns

Gambar 3 Hasil Scraping

3.2. Pelabelan Otomatis

Setelah pengumpulan data peneliti selanjutnya melakukan pelabelan pada aplikasi shopee. Berikut adalah hasil pelabelan dengan script yang tertera pada gambar.

Pelabelan Data

```
def pelabelan (rate):
    if rate <=3:
        return 'negatif'
    elif rate >=4:
        return 'positif'

data ['label'] = data ['score'].apply(pelabelan)
data
```

Berikut hasil outputnya.

	userName	score	at	content	label
0	revina Safitri	2	2022-11-25 06:48:16	Sumpah lemot banget please, sering nge bug git...	negatif
1	Auliarubi Kusmawan	5	2022-11-25 06:47:52	Bagusssssssssssssssssssssssssssssssss	positif
2	John Bay	5	2022-11-25 06:47:47	Joshh	positif
3	Davit Irawan	5	2022-11-25 06:46:22	Mantap 🤩	positif
4	rina wahyu	5	2022-11-25 06:46:08	Shopee kok sering macet ya, jd harus sering tu...	positif
...	...	...	...	...	...
995	Nabila Zulaikha	1	2022-11-24 12:08:56	APA SIH MAKIN HARI SHOPEE UDAH ENGGAK KAYA DUK...	negatif
996	Mohamad Silva Sulhana	3	2022-11-24 12:08:55	ongkirnya lbh mahal drpd barang yg diberi ya, ...	negatif
997	Syafira Qiandra	1	2022-11-24 12:08:05	Duh kenapa ada biaya macam macam sekarang apal...	negatif
998	willy kris Hadiyanto	5	2022-11-24 12:07:52	Mudah dan keren abis	positif
999	Suyatmi Sby	5	2022-11-24 12:07:22	Memuaskan	positif

1000 rows x 5 columns

Gambar 4 Hasil Pelabelan

Gambar 4 adalah contoh data yang telah diberi label, Jika sebelumnya hanya terdapat kolom user, score, at dan content. Sehingga akan terbentuk tabel tambahan yang disebut sentiment, yang digunakan untuk mengetahui sifat dari teks yang didapat.

3.3. Preprocessing

Selanjutnya peneliti melakukan tahap *preprocessing*. berikut adalah hasil pengambilan sampel teks untuk dilakukan *preprocessing* dengan mengambil satu sampel text. Diawali dengan data asli dari hasil proses *scraping*, dilanjutkan dengan proses *case folding*, *tokenizing*, *filtering* dan *stemming*.

Table 1 Preprocessing

Preprocessing	Content
Data Mentah	Sumpah lemot banget please, sering nge bug gituu...inih bagian back end nya makan gaji butah kah??
Case Folding	sumpah lot banget please nge bug gituuinih back end nya makan gaji butah kah
Tokenize	[sumpah,lemot,banget,please,sering,nge,bug,gituuinih,bagian,back,end,nya,makan,gaji,butah,kah]
Filtering	[sumpah,lemot,banget,please,nge,bug,gituuinih,back,end,nya,makan,gaji,butah,kah]
Stemming	sumpah lot banget please nge bug gituuinih back end nya makan gaji butah kah

### 3.4. Hasil Accuracy

Pada tahap klasifikasi yang dilakukan dengan cara membuat sebuah *machine learning* menggunakan data training dan data testing pada seluruh data secara random pada dataset untuk melakukan cross validation dan menghasilkan nilai prediksi untuk akurasi. Dibawah ini merupakan gambaran dari hasil tahapan klasifikasi menggunakan *script* dari algoritma KNN.

#### Algoritma KNN

```

from sklearn.neighbors import KNeighborsClassifier
classifier = KNeighborsClassifier(n_neighbors= 3)
classifier.fit(x_train_cv, y_train)
y_pred = classifier.predict(x_test_cv)
from sklearn.metrics import classification_report, confusion_matrix
cm = confusion_matrix(y_test, y_pred)
cr = classification_report(y_test, y_pred)
print(cm)
print(cr)

```

Output dari *script* diatas dapat terlihat pada gambar dibawah ini.

[[33 67]					
[24 76]]					
	precision	recall	f1-score	support	
negatif	0.58	0.33	0.42	100	
positif	0.53	0.76	0.63	100	
accuracy			0.55	200	
macro avg	0.56	0.55	0.52	200	
weighted avg	0.56	0.55	0.52	200	

Gambar 5 Hasil Akurasi KNN

Gambar 5 merupakan hasil Validasi dari penerapan algoritma K-Nearest Neighbor pada aplikasi shopee menunjukkan bahwa tingkat akurasi 55,00%. dimana jumlah data untuk sentimen negatif sebanyak 100 ulasan dan sentimen positif sebanyak 100 ulasan dari data testing yang digunakan sebanyak 200 data atau 20% dari jumlah data yang digunakan dengan metode random pada saat testing.

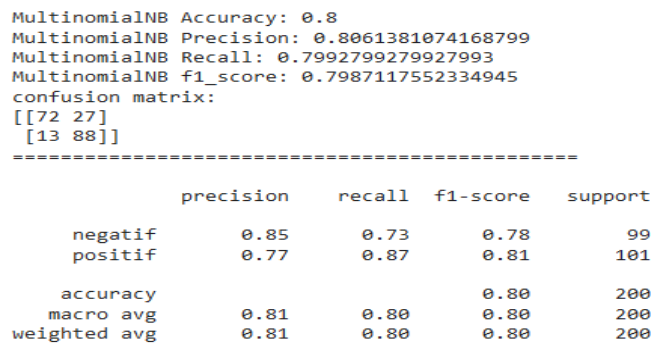
Kemudian setelah hasil dari KNN telah keluar dilanjutkan dengan metode berikutnya menggunakan script dari algoritman NBC berikut script dan hasilnya.

Algoritma NBC

```

from sklearn.naive_bayes import MultinomialNB
from sklearn.metrics import accuracy_score, precision_score, recall_score, f1_score
from sklearn.metrics import classification_report
from sklearn.metrics import confusion_matrix
clf = MultinomialNB().fit(x_train, y_train)
predicted = clf.predict(x_test)
print("MultinomialNB Accuracy:", accuracy_score(y_test, predicted))
print("MultinomialNB Precision:", precision_score(y_test, predicted, average="macro", pos_label="negatif"))
print("MultinomialNB Recall:", recall_score(y_test, predicted, average="macro", pos_label="negatif"))
print("MultinomialNB f1_score:", f1_score(y_test, predicted, average="macro", pos_label="negatif"))
print(f'confusion matrix:\n{confusion_matrix(y_test, predicted)}')
print('=====')
print(classification_report(y_test, predicted, zero_division=0))
    
```

Output script diatas dapat dilihat pada gambar dibawah ini.



Gambar 6 Hasil Akuasi NBC

Pada gambar 6, merupakan hasil Validasi dari penerapan algoritma Naïve Bayes Classifier pada aplikasi shopee menunjukkan bahwa tingkat akurasi 80,00%. dimana jumlah data untuk sentimen negatif sebanyak 99 ulasan dan sentimen positif sebanyak 101 ulasan dari data testing yang digunakan sebanyak 200 data atau 20% dari jumlah data yang digunakan dengan metode random pada saat testing.

Pada penelitian ini perbandingan kinerja algoritma dapat dilihat dari hasil akurasi pada tabel 3 sebagai berikut:

Table 2 Hasil Akurasi

NO	Algoritma	Akurasi
1	Naïve Bayes Classifier	80,00 %
2	K-Nearst Neighbor	55,00%

Berdasarkan pemaparan perbandingan kinerja terlihat bahwa hasil akurasi Algoritma Naïve Bayes Classifier dengan akurasi sebesar 80,00% dan Algoritma K-Nearst Neighbor dengan akurasi sebesar 55,00%. Artinya akurasi Algoritma Naïve Bayes Classifier lebih besar dari Algoritma K-Nearst Neighbor, sehingga bisa ditarik kesimpulan bahwan Algoritma Naïve Bayes Classifier mempunyai kinerja yang baik dalam klasifikasi.

4. KESIMPULAN

Tujuan dari penelitian ini adalah untuk mengetahui tingkat akurasi dari penggunaan dua metode yang berbeda antara lain metode K-Nearst Neighbor dan metode Naïve Bayes Classifier yang dibantu dengan penggunaan *tools colab google*. Pada proses ini peneliti diharuskan untuk melakukan pencarian dan pengumpulan data pada ulasan komentar yang telah tersedia pada *google play store* yang selanjutnya dilakukan proses pengolahan data diantaranya filtering, labeling, yang dilanjutkan dengan analisa agar dapat mengetahui tingkat akurasi pada kedua metode tersebut. Pada penelitian kali ini bisa diketahui untuk metode K-Nearst Neighbor dan Naïve Bayes Classifier mempuni untuk dipergunakan dalam pemrosesan dibidang data *mining*. Analisis sentimen yang dilakukan pada data yang tersedia di *play store* perihal aplikasi shopee medapatkan skor akurasi 55.00% dengan jumlah data yang digunakan untuk sentimen negatif sebanyak 100 ulasan dan sentimen positif sebanyak 100 ulasan dari data testing yang digunakan sebanyak 200 data atau 20% dalam metode K-Nearst Neighbor. Lalu tingkat akurasi mencapai 80,00% dimana jumlah data untuk sentimen negatif sebanyak 99 ulasan dan sentimen positif sebanyak 101 ulasan dari data testing yang digunakan sebanyak 200 data atau 20% pada metode Naïve Bayes Classifier.

Untuk penelitian yang telah dilakukan peneliti, bisa dilihat bahwa metode Naïve Bayes Classifier merupakan metode dengan tingkat akurasi yang lebih tinggi dibandingkan K-Nearst Neighbor dengan tingkat akurasi sebesar 80,00%.

**DAFTAR PUSTAKA**

- [1] E. S. Sulistiyawati and A. Widayani, "Marketplace Shopee Sebagai Media Promosi Penjualan UMKM di Kota Blitar," *Jurnal Pemasaran Kompetitif*, vol. 4, no. 1, p. 133, Oct. 2020, doi: 10.32493/jpkpk.v4i1.7087.
- [2] P. Studi Manajemen, K. Ilmiyah, I. Krishernawan, S. Al-Anwar Jalan Raya Brangkal No, S. Mojokerto, and J. Timur, "Maker: Jurnal Manajemen MARKETPLACE SHOPEE DI MOJOKERTO," 2020, [Online]. Available: <http://www.maker.ac.id/index.php/maker>
- [3] G. Pratama, "Analisis Transaksi Jual Beli online Melalui Website Marketplace Shopee Menurut Konsep Bisnis di Masa Pandemic Covid 19," 2020. [Online]. Available: <http://journal.bungabangsacirebon.ac.id/index.php/ecopreneur>
- [4] R. Azhar, A. Surahman, and C. Juliane, "Analisis Sentimen Terhadap Cryptocurrency Berbasis Python TextBlob Menggunakan Algoritma Naïve Bayes," 2022.
- [5] O. Nurdiawan, R. Herdiana, and S. Anwar, "Komparasi Algoritma Naïve Bayes dan Algoritma K-Nearest Neighbor terhadap Evaluasi Pembelajaran Daring," *SMATIKA JURNAL*, vol. 11, no. 02, pp. 126–135, Dec. 2021, doi: 10.32664/smatika.v11i02.621.
- [6] H. Putri, A. I. Purnamasari, A. R. Dikananda, O. Nurdiawan, and S. Anwar, "Penerima Manfaat Bantuan Non Tunai Kartu Keluarga Sejahtera Menggunakan Metode NAÏVE BAYES dan KNN," *Building of Informatics, Technology and Science (BITS)*, vol. 3, no. 3, pp. 331–337, Dec. 2021, doi: 10.47065/bits.v3i3.1093.
- [7] J. Homepage, N. C. Agustina, D. Herlina Citra, W. Purnama, C. Nisa, and A. Rozi Kurnia, "MALCOM: Indonesian Journal of Machine Learning and Computer Science The Implementation of Naïve Bayes Algorithm for Sentiment Analysis of Shopee Reviews on Google Play Store Implementasi Algoritma Naïve Bayes untuk Analisis Sentimen Ulasan Shopee pada Google Play Store," vol. 2, pp. 47–54, 2022.
- [8] M. I. Fikri, T. S. Sabrila, Y. Azhar, and U. M. Malang, "Perbandingan Metode Naïve Bayes dan Support Vector Machine pada Analisis Sentimen Twitter"
- [9] W. E. Nurjanah, R. Setya Perdana, and M. A. Fauzi, "Analisis Sentimen Terhadap Tayangan Televisi Berdasarkan Opini Masyarakat pada Media Sosial Twitter menggunakan Metode K-Nearest Neighbor dan Pembobotan Jumlah Retweet," 2017. [Online]. Available: <http://j-ptiik.ub.ac.id>
- [10] R. Gunawan, R. Septiadi, F. Apri Wenando, H. Mukhtar, and Syahril, "K-Nearest Neighbor (KNN) untuk Menganalisis Sentimen terhadap Kebijakan Merdeka Belajar Kampus Merdeka pada Komentar Twitter," *Jurnal CoSciTech (Computer Science and Information Technology)*, vol. 3, no. 2, pp. 152–158, Aug. 2022, doi: 10.37859/coscitech.v3i2.3841.
- [11] A. P. Giovani, A. Ardiansyah, T. Haryanti, L. Kurniawati, and W. Gata, "ANALISIS SENTIMEN APLIKASI RUANG GURU DI TWITTER MENGGUNAKAN ALGORITMA KLASIFIKASI," *Jurnal Teknoinfo*, vol. 14, no. 2, p. 115, Jul. 2020, doi: 10.33365/jti.v14i2.679.
- [12] N. Kusuma Wardhani<sup>1</sup> *et al.*, "SENTIMENT ANALYSIS ARTICLE NEWS COORDINATOR MINISTER OF MARITIME AFFAIRS USING ALGORITHM NAIVE BAYES AND SUPPORT VECTOR MACHINE WITH PARTICLE SWARM OPTIMIZATION," *J Theor Appl Inf Technol*, vol. 31, p. 24, 2018, [Online]. Available: [www.jatit.org](http://www.jatit.org)
- [13] E. B. Susanto, Paminto Agung Christianto, Mohammad Reza Maulana, and Satriedi Wahyu Binabar, "Analisis Kinerja Algoritma Naïve Bayes Pada Dataset Sentimen Masyarakat Aplikasi NEWSAKPOLE Samsat Jawa Tengah," *Jurnal CoSciTech (Computer Science and Information Technology)*, vol. 3, no. 3, pp. 234–241, Dec. 2022, doi: 10.37859/coscitech.v3i3.4343.