

ANALISIS SENTIMEN REVIEW PELANGGAN MARKETPLACE SHOPEE INDONESIA MENGGUNAKAN METODE ALGORITMA K-NEAREST NEIGHBORS

Leni Susanti^{1,*}, Akrom²

¹Program Studi Sistem Informasi Universitas Pamulang

²Program Studi Teknik Informatika Universitas Pamulang

Jl. Raya Puspitek No.46, Buaran, Serpong

*E-mail: dosen02617@unpam.ac.id

ABSTRAK

ANALISIS SENTIMEN REVIEW PELANGGAN MARKETPLACE SHOPEE INDONESIA MENGGUNAKAN METODE ALGORITMA K-NEAREST NEIGHBORS. Berbelanja secara *on-line* sekarang merupakan aktivitas rutin yang terjadi pada masyarakat. Berkembangnya jaman membuat seseorang memilih untuk belanja secara *on-line* daripada harus melakukan perjalanan ke toko untuk mendapatkan kebutuhan yang diinginkan. Salah satu marketplace yang saat ini berkembang dengan pesat adalah Shopee Indonesia. Shopee Indonesia ditujukan sebagai salah satu pusat perbelanjaan yang dikelola oleh Garena (berubah nama menjadi SEA Group). Bisnis *Customer to Customer (C2C) mobile marketplace* yang digunakan oleh Shopee memungkinkan kehadirannya dengan mudah diterima oleh semua lapisan masyarakat, termasuk di Indonesia. Semenjak awal peluncurannya, Shopee Indonesia mengalami peningkatan yang sangat pesat, bahkan hingga saat ini aplikasi Shopee menempati urutan kedua baik dari segi pengguna aktif bulanan maupun jumlah download aplikasi Shopee di Indonesia. Sebuah *review* barang menjadi penting pada setiap barang di sebuah toko *on-line*, karena hal ini sangat berguna untuk melihat bagaimana umpan balik pembeli sebelumnya melalui sebuah komentar. Komentar positif atau komentar negatif merupakan hal wajar yang terjadi dalam proses *review* produk. Untuk mengatasi masalah tersebut maka digunakan metode analisis sentimen *review* produk pada aplikasi Shopee, dimana salah satu algoritma untuk mengklasifikasikan komentar positif atau negatif yaitu algoritma *K-Nearest Neighbor (K-NN)*. Pemilihan metode *K-Nearest Neighbor (K-NN)* dikarenakan memiliki kelebihan untuk mencapai hasil akurasi rule baik dan sesuai dengan perhitungan berdasarkan aturan yang diterapkan dalam sebuah aplikasi. Dalam penelitian ini menggunakan 150 komentar dimana terbagi menjadi dua yaitu 75 komentar positif dan 75 komentar negatif. Di harapkan dengan penelitian ini dapat mengetahui seberapa besar nilai akurasi yang dihasilkan dengan proses klasifikasi dengan metode *K-Nearest Neighbor (K-NN)*. Hasil dari penelitian ini adalah dengan mendapatkan nilai akurasi dengan menggunakan *K-Nearest Neighbor (KNN)* adalah sebesar 66.67 %, nilai *precision* positif sebesar 64,71 %, nilai *precision* negatif sebesar 69,23 %, nilai *recall* positif sebesar 73.33 % dan nilai *recall* negatif sebesar 60 % . Sedangkan Luaran dari penelitian ini adalah jurnal

Kata kunci: Analisis Sentimen, Review Produk, Aplikasi Shopee, Klasifikasi, *K-Nearest Neighbor (K-NN)*

ABSTRACT

ANALYSIS OF SHOPEE INDONESIA CUSTOMER REVIEW SENTIMENT USING THE K-NEAREST NEIGHBORS ALGORITHM. Online shopping is now a routine activity that occurs in society. The development of the era makes a person choose to shop online instead of having to travel to the store to get the desired needs. One marketplace that is currently growing rapidly is Shopee Indonesia. Shopee Indonesia is intended as a shopping center managed by Garena (changed name to SEA Group). The Customer to Customer (C2C) mobile marketplace business used by Shopee allows its presence to be easily accepted by all levels of society, including in Indonesia. Since its initial launch, Shopee Indonesia has experienced a very rapid increase, even today the Shopee application ranks second both in terms of monthly active users and the number of downloads of the Shopee application in Indonesia. An item review is important for every item in an on-line store, because it is very useful to see how previous buyers feedback through a comment. Positive comments or negative comments are normal things that happen in the product review process. To overcome this problem, the product review sentiment analysis method is used in the Shopee application, where one of the algorithms for classifying positive or negative comments is the *K-Nearest Neighbor (K-NN)* algorithm. The choice of the *K-Nearest Neighbor (K-NN)* method is because it has the advantage of achieving good rule accuracy results and in accordance with calculations based on the rules applied in an application. In this study, 150 comments were used which were divided into two, namely 75 positive comments and 75 negative comments. It is hoped that this research can find out how much the accuracy value is generated by the classification

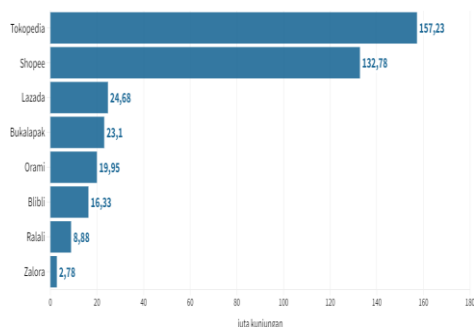
process with the *K-Nearest Neighbor (K-NN)* method. The results of this study are to obtain an accuracy value using the *K-Nearest Neighbor (KNN)* of 66.67%, a positive precision value of 64.71%, a negative precision value of 69.23%, a positive recall value of 73.33% and a recall value negative by 60%. While the output of this research is a journal

Keywords: *Sentiment Analysis, Product Review, Shopee, Classification, K-Nearest Neighbor (K-NN)*

1. PENDAHULUAN

Teknologi dan informasi khususnya internet saat ini mengalami perkembangan yang cukup pesat serta dalam perkembangannya telah mengubah berbagai aspek kehidupan manusia, hal ini dikarenakan internet telah menjadi suatu kebutuhan pokok selain kebutuhan pangan, sandang, dan papan bagi sebagian besar orang terutama di daerah perkotaan (Ho & Lee, 2007). Salah satu manfaat dari media internet dalam aplikasi bisnis yaitu marketplace atau sering juga disebut perdagangan elektronik. Marketplace merupakan perdagangan yang memanfaatkan internet untuk membeli, menjual, mengangkut, atau memperdagangkan data, barang, atau layanan (Turban, 2015).

Marketplace dapat diartikan sebagai proses pembelian dan penjualan produk yang memanfaatkan elektronik data melalui internet (Grandon & Pearson, 2004). Di Indonesia sendiri, pertumbuhan dan perkembangan untuk marketplace semakin meningkat. Salah satu perusahaan marketplace yang sukses dan berhasil dalam memanfaatkan peluang pasar marketplace di Indonesia yaitu Shopee. Shopee Indonesia ditunjukan sebagai salah satu pusat perbelanjaan yang dikelola oleh Garena (berubah nama menjadi SEA Group). Semenjak awal peluncurannya, Shopee Indonesia mengalami peningkatan yang sangat pesat, bahkan hingga saat ini aplikasi Shopee menempati urutan kedua baik dari segi pengguna aktif bulanan maupun jumlah download aplikasi Shopee di Indonesia.



Gambar 1. Perangkingan E-Commerce (sumber situs iPrice)

Bagian terpenting dari suatu *sentiment analysis* adalah *opinion mining*, *opinion mining* merupakan proses memahami, mengekstrak serta mengolah data tekstual secara otomatis untuk mendapatkan informasi yang ingin dicapai. Perlu dilakukan analisis sentimen untuk melihat pendapat atau kecenderungan opini terhadap sebuah masalah atau objek oleh seseorang, hal ini untuk menunjukkan apakah opini tersebut cenderung berpandangan negatif atau positif (Bo & Lee, 2008).

Terdapat beberapa kelebihan yang dimiliki oleh metode *K-Nearest Neighbor (K-NN)* yaitu algoritma klasifikasi *K-Nearest Neighbor* terbukti dapat mencapai hasil akurasi yang baik dan sesuai dengan perhitungan yang akan diterapkan dalam sebuah aplikasi (Sani, Zeniarza, & Luthfiarta, 2016). Berdasarkan uraian latar belakang dan identifikasi masalah diatas, maka diperlukan penelitian mengenai analisis sentimen *review* pengguna terhadap aplikasi Shopee Indonesia dengan memanfaatkan kinerja algoritma *K-Nearest Neighbor (K-NN)* untuk klasifikasi *review* terhadap komentar negatif dan positif. Diharapkan pada penelitian ini bisa membantu pengguna aplikasi Shopee baik dari pihak konsumen maupun produsen dalam memahami *review* kekurangan serta kelebihan dari aplikasi Shopee itu sendiri.

2. TINJAUAN PUSTAKA

Membahas tentang teori dan penelitian terdahulu yang berkaitan dengan topik penelitian yang menjadi landasan logis dalam mengembangkan hipotesis penelitian termasuk kerangka konsep penelitian.

A. Data Mining

Data mining adalah proses untuk menemukan pola yang berguna dan kecenderungan di dalam kumpulan data yang besar (Larose & Larose, 2015). Sebuah variasi ada dalam hal masalah *domain*, aplikasi, representasi formulasi, dan data yang ditemui dalam aplikasi nyata. Oleh karena itu, "*data mining*" adalah istilah payung yang luas yang digunakan untuk menggambarkan aspek-aspek yang berbeda dari pengolahan data.

B. Text Mining

Text Mining dapat juga diartikan sebagai penambangan data berupa teks yang bersumber dari dokumen untuk mencari kata-kata yang merupakan perwakilan isi atau pembentuk dokumen teks sehingga penganalisisan dapat dibuat. Berikut ini merupakan tahapan umum pada proses *Text Mining* yaitu (Feldman & Sanger, 2007):

1. *Text Preprocessing*
2. Case Folding
3. Filtering
4. Tokenizing
5. Stemming

C. K-Nearest Neighbour (KNN)

Algoritma *K-Nearest Neighbor* (K-NN) adalah sebuah metode untuk melakukan klasifikasi terhadap objek berdasarkan data pembelajaran yang jaraknya paling dekat dengan objek tersebut (Widiarsana, O et al., 2011). Ketepatan algoritma *K-Nearest Neighbor* sangat dipengaruhi oleh ada atau tidaknya fitur-fitur yang tidak relevan atau jika bobot fitur tersebut tidak setara dengan relevansinya terhadap klasifikasi. Menurut Kusriani dan Emha (2009) algoritma *K-Nearest Neighbor* adalah pendekatan untuk mencari kasus dengan menghitung kedekatan antara kasus baru dengan kasus lama dengan berdasarkan pada pencocokan bobot dari sejumlah fitur yang ada yang memiliki kesamaan (*similarity*).

3. METODE

Data yang digunakan pada penelitian ini merupakan data sentimen dari pengguna aplikasi Shopee yang memberikan komentarnya pada Google Playstore yang kemudian dipisahkan menjadi sentimen positif dan sentimen negatif. Sentimen positif yaitu sentimen dimana pengguna memberikan komentarnya dengan memberikan penilaian secara positif atau memberikan keterangan bahwa aplikasi ini baik atau mempunyai kepuasan dalam penggunaan aplikasi Shopee Indonesia tersebut. Sedangkan sentimen negatif merupakan sentimen sebaliknya dari sentimen positif yaitu pengguna memberikan komentarnya dengan memberikan ketidakpuasan terhadap aplikasi Shopee Indonesia. Berikut dijelaskan tahapan-tahapan dari pengambilan data:

- a. Buka Aplikasi Google Playstore, pada tahap ini dilakukan proses daftar dengan email untuk dapat menggunakan aplikasi Google Playstore.
- b. Cari aplikasi Shopee Indonesia tentu saja ada *tools* pencarian aplikasi yang dijadikan sampel dalam penelitian ini.
- c. Cek komentar para pengguna aplikasi Shopee Indonesia, melakukan observasi terhadap isi komentar/*review* dari penilaian aplikasi dari pengguna Shopee.
- d. *Save Data*, menyimpan data *review* atau penilaian produk tersebut.

Tabel 1. Contoh Sampel Data Mentah Sentimen.

X	create_at	score	Text
1	3-May-23 21:41	5	Bagus.. Suka pake sopi. Kalo lemot biasanya minta diupdate. Sebulan sekali atau 1 bulan 2x mesti minta diupdate. Kalo udah diupdate bisa lancar. Sangat membantu untuk belanja.
2	26-Apr-23 15:20	2	Ada menu OneKlik BCA, tapi gak bisa digunakan dengan keterangan gagal dan disuruh menggunakan metode lain, padahal ada menunya dan lebih simple klo bisa pake Oneklik, ayo tim pengembang aplikasi, segera di perbaiki lagi bug nya.
3	26-Apr-23 12:48	1	Makin sering di update shopee saya gak bisa COD lagi kualitasnya perbaiki lagi dah bagus yang dulu dari pada yang sekarang, trss pesanan sudah di batalkan ada keterangan nya ttp aja pakatnya datang. Mohon shopee diperhatikan lagi dong!
4	20-Mar-23 02:01	4	Jual beli aman, harga brsaing, murah. Namun sayang ada beberapa fitur jika mo merubah pembayaran tidak bisa, jadi harus diulang dr awal (dibatalkan dl lalu pesan ulang). Aplikasi berat, loading lama, tidak seperti lapak sebelah..cb tim IT nya suruh cek lg apa yg bikin berat, kadang ngeselin lama loadingnya. Trims.

4. HASIL DAN PEMBAHASAN

4.1 Hasil

Proses pengambilan data dilakukan dengan cara menggunakan bantuan ekstensi Google Chrome yaitu dengan menggunakan *Data Scraper*. Penggunaan ekstensi *Data Scraper* merupakan salah satu ekstensi pada aplikasi Google Chrome yang dapat dimanfaatkan sebagai web *scraper* dalam proses pengambilan data komentar atau konten yang terdapat pada situs online secara gratis. Data yang diekstraksi dengan menggunakan *Data Scraper* dimana data dihasilkan dari halaman web HTML dapat diimpor ke dalam bentuk spreadsheet Microsoft Excel. Proses pengambilan dengan data scrapper dilakukan dengan cara login ke dalam akun *Data Scraper*, selanjutnya adalah dengan melakukan proses pengambilan data dari aplikasi Shopee yang telah ditentukan Colab <https://colab.research.google.com/> dengan script berikut kemudian klik run untuk diambil komentarnya dari para pengguna dengan bantuan menggunakan Google. Berikut adalah contoh proses pengambilan data.

```

[1] #Referensi: https://www.linkedin.com/pulse/how-scrape-google-play-reviews-4-simple-steps-using-python-kundi/
#download library google-play-scraper
!pip install google-play-scraper

Looking in indexes: https://pypi.org/simple, https://us-python.org:https://api.python.org/dev/colab-nheels/public/simple/
Collecting google-play-scraper
  Downloading google_play_scraper-1.2.4-py3-none-any.whl (28 kb)
Installing collected packages: google-play-scraper
Successfully installed google-play-scraper-1.2.4

[2] from google_play_scraper import app
import pandas as pd
import numpy as np

[3] #scrape jumlah ulasan yang diinginkan
from google_play_scraper import Sort, reviews
result, continuation_token = reviews(
    'com.shopee.id',
    lang='id', #defaults to 'en'
    country='id', #bahasa yang digunakan adalah bahasa Indonesia
    sort=Sort.MOST_RELEVANT, # kemudian gunakan most_relevant untuk mendapatkan ulasan yang paling relevan
    count=150, # merupakan jumlah ulasan yang ingin diambil sebagai sampel
    filter_score_min=None #filter_score menggunakan None untuk mengambil semua atau rating dari bintang 1 sampai 5
)

```

Gambar 2. Tahap *Data Scrapper*

Data yang tersimpan dari proses *scraping data* selanjutnya dilakukan proses labelling untuk membedakan antara komentar positif dan komentar negatif. Dikarenakan pada penelitian berikut menggunakan 150 komentar yang terdiri dari 75 komentar positif dan komentar negatif, maka proses pelabelan dilakukan dengan cara memilah jenis komentar berdasarkan *score*. Komentar dengan *score* 4 dan 5 diberikan label sebagai komentar positif sedangkan untuk komentar dengan *score* 1, 2 dan 3 diberikan label sebagai komentar negatif. Untuk mempermudah proses klasifikasi pada sistem maka komentar positif ditandai dengan label 0 sedangkan untuk komentar negatif ditandai dengan label 1. Adapun

penggunaan code dalam proses pelabelan adalah sebagai berikut:

- a. Case Folding
- b. Stopward Removal
- c. Tokenizing
- d. Stemming

Hasil analisa dalam bentuk *wordcloud* untuk komentar positif ditunjukkan oleh visualiasasi gambar 3, sedangkan komentar negatif ditunjukkan pada gambar 4.



Gambar 3. Visualisasi Komentar Positif dengan *Wordcloud*.

Berdasarkan hasil visualisasi pada gambar 3 di atas, terlihat bahwa kata pada komentar positif yang sering muncul pada komentar yang diberikan oleh pengguna aplikasi shopee adalah kata yang dicetak lebih besar daripada kata yang lain. Dimana dari *wordcloud* dapat disimpulkan bahwa pengguna lebih sering memberikan komentar positif dengan kata “barang” pada komentar yang diberikan pada aplikasi Shopee di Google Playstore. Selain kata “barang” pengguna juga memberikan komentar positif dengan kata “belanja”, “bagus”, “baik”, dan “ongkir”.



Gambar 4. Visualisasi Komentar Negatif dengan *Wordcloud*.

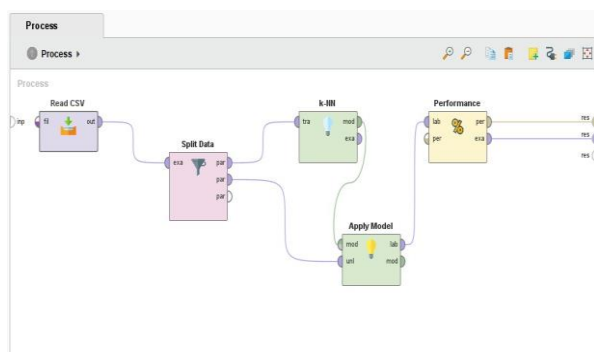
Sedangkan berdasarkan hasil visualisasi pada gambar 4. tentang komentar negatif, terlihat bahwa kata pada komentar negatif yang sering muncul pada komentar yang diberikan oleh

pengguna aplikasi shopee adalah kata “berat” pada komentar yang diberikan pada aplikasi Shopee di Google Playstore. Selain kata “berat” pengguna juga memberikan komentar negatif dengan kata “lemot”, “loading”, “parah”, dan “fitur”.

4.2 Pembahasan

1. Penerapan Klasifikasi *K-Nearest Neighbor* (KNN)

Setelah dilakukan tahapan pembobotan dari masing-masing *term* dan *vector* dari masing-masing dokumen, maka akan dilakukan klasifikasi menggunakan metode *K-Nearest Neighbor* (K-NN). Pada proses penerapan perhitungan dataset dengan menggunakan *K-Nearest Neighbor* (KNN) dilakukan dengan menggunakan *tools Rapidminer Studio*. Hal pertama dalam penggunaan metode *K-Nearest Neighbor* (KNN) adalah menentukan jumlah kelas. Disini peneliti menentukan nilai $k = 3$. Adapun penerapan dalam proses klasifikasi *K-Nearest Neighbor* (K-NN) dengan menggunakan RapidMiner ditunjukkan pada gambar berikut:



Gambar 5. Penerapan Klasifikasi *K-Nearest Neighbor* (KNN) dengan RapidMiner

2. Uji dan Evaluasi Model

Untuk mengetahui performance dari algoritma *K-Nearest Neighbor* (KNN), maka perlu dilakukan pengujian sebuah model. Pada pengujian model dengan menggunakan *tools Rapidminer* yang telah dilakukan pada Gambar 5, hasil klasifikasi akan ditampilkan dalam bentuk *confusion matriks*. Tabel *confusion matriks* terdiri atas kelas *predicted* dan kelas *actual*. Model *confusion matriks* 2x2 ditunjukkan tabel sebagai berikut:

Tabel 2. Model *Confusion Matriks*

<i>Predicted</i>	<i>Actual</i>	
	Positif	Negatif
Positif	11	6
Negatif	4	9

Untuk mengetahui nilai *accuracy*, *precision* dan *recall* dapat dilakukan perhitungan secara manual sebagai berikut:

$$Accuracy = \frac{11 + 9}{11 + 4 + 6 + 9} = 0.66$$

$$Precision = \frac{11}{11 + 6} = 0.64$$

$$Recall = \frac{11}{11 + 4} = 0.73$$

Dengan menggunakan bantuan rapidminer *accuracy*, *precision* dan *recall* maka didapatkan dengan keterangan sebagai berikut:

accuracy: 66.67%

	true Positif	true Negatif	class precision
pred Positif	11	6	64.71%
pred Negatif	4	9	69.23%
class recall	73.33%	60.00%	

Gambar 6. Nilai *accuracy*, *precision* dan *recall*

Dapat disimpulkan bahwa sentiment analisis pada 150 data komentar yang terbagi menjadi 75 komentar negatif dan 75 komentar positif jika dilakukan dengan menggunakan metode *K-Nearest Neighbor* (KNN) akan menghasilkan nilai akurasi sebesar 66,67 %.

5. KESIMPULAN

Dari hasil analisis yang sudah dilakukan dengan menggunakan *dataset* yang diambil dari kolom komentar aplikasi Shopee Indonesia dapat diambil beberapa kesimpulan sebagai berikut:

- Nilai Akurasi yang didapatkan dengan menggunakan *K-Nearest Neighbor* (KNN) adalah sebesar 66.67 %, nilai precision positif sebesar 64,71 %, nilai precision negatif sebesar 69,23 %, nilai recall positif sebesar 73.33 % dan nilai recall negatif sebesar 60 %

- b. Dari hasil analisa dengan menggunakan *wordcloud* dapat disimpulkan bahwa pengguna dalam komentar positifnya yang sering digunakan adalah kata “belanja”, “bagus”, “baik”, dan “ongkir”. Sedangkan untuk komentar negatifnya yang sering digunakan adalah kata “lemot”, “loading”, “parah”, dan “fitur”. Hal ini dapat membantu pengambilan keputusan untuk memperbaiki aplikasi oleh para management Shopee.

DAFTAR PUSTAKA

- Cui, H. e. (2008). Comparative Experiments on Sentiment Classification for Online Product Reviews. Singapore: Department of Computer Science , National University of Singapore.
- Grandon, E., & Pearson, M. (2004). E-Commerce adoption: Perceptions of Managers/Owners of Small and Medium Enterprises in Chile. *Communication of The Association for Information Systems* 13, 81-102.
- Ho, C.-I., & Lee, Y.-L. (2007). The development of an e-travel service quality scale. *Tourism Management* Volume 28 No.6, 1434-1449.
- Kusrini, & Emha, T. L. (2009). *Algoritma Data Mining*. Yogyakarta: Andi.
- Sani, R., Zeniarza, J., & Luthfiarta, A. (2016). Pengembangan Aplikasi Penentuan Tema Tugas Akhir Berdasarkan Data Abstrak Menggunakan Algoritma K-Nearest Neighbor. *Prosiding seminar Nasional Multi Disiplin Ilmu & Call For Papers*, 103-111.
- Samuel, Y., Delima, R., & Rachmat, A. (2014). Implementasi Metode K-Nearest Neighbor dengan Decision Rule untuk Klasifikasi Subtopik Berita. *Jurnal Informatika* Volume 10 No.1,, 1-15.
- Turban, E. (2015). *Electronic commerce a Managerial and Social Networks Perspective* 8th edition. Springer.