

Analisis Sentimen Terhadap Penggunaan Aplikasi Shopee Menggunakan Algoritma *Support Vector Machine* (SVM)

Irma Surya Kumala Idris
Program Studi Teknik Informatika
Universitas Ichsan Gorontalo
Gorontalo, Indonesia
mhaladp@gmail.com

Yasin Aril Mustofa*
Program Studi Teknik Informatika
Universitas Ichsan Gorontalo
Gorontalo, Indonesia
arielddcc@gmail.com

Irvan Abraham Salihi
Program Studi Teknik Informatika
Universitas Ichsan Gorontalo
Gorontalo, Indonesia
irvanabrahams@gmail.com

Diterima : Agustus 2022
Disetujui : Oktober 2022
Dipublikasi : Januari 2023

Abstrak—Analisis Sentimen merupakan cabang dari penelitian text mining yang melakukan proses pengklasifikasian dokumen teks. Analisis sentimen dapat melakukan ekstraksi pendapat, emosi, dan evaluasi tertulis seseorang tentang topik tertentu menggunakan teknik pemrosesan Bahasa alami. Pada penelitian ini melakukan analisis sentiment terhadap penggunaan aplikasi Shopee menggunakan algoritma *Support Vector Machine* (SVM). Tujuan dari penelitian ini adalah untuk mengklasifikasi data komentar dari pengguna aplikasi Shopee kedalam komentar positif dan negatif dengan mempelajari pendapat pengguna tentang aplikasi Shopee melalui ulasan yang diberikan, dan untuk mengetahui kinerja dari metode pengklasifikasi yang digunakan. Pada penelitian ini data diperoleh dengan cara mengangkat data dari ulasan penggunaan aplikasi Shopee menggunakan metode *scraping* dan berhasil mendapat 3000 data ulasan. Hasil penelitian menggunakan algoritma *Support Vector Machine* terbukti mampu menghasilkan kinerja yang cukup baik dengan hasil akurasi sebesar 98% dan f1-score sebesar 0.98 atau sebesar 98%.

Kata Kunci—SVM; Shopee; Analisis Sentimen

Abstrak—Sentiment analysis is a branch of text mining research that carries out the process of classifying text documents. Sentiment analysis can extract one's opinions, emotions, and evaluations about a certain topic using natural language techniques. In this study, sentiment analysis was carried out on the use of the Shopee application using the Support Vector Machine (SVM) algorithm. The purpose of this study is to classify comment data from Shopee application users, positive and negative comments by studying user opinions about the Shopee application through the reviews provided, and to determine the performance of the classifier method used. In this study, the data was obtained by collecting data from reviews on the use of the Shopee application using the scraping method and managed to get 3000 data reviews. The results of research using the Support Vector Machine algorithm are proven to be able to produce quite good performance with an accuracy of 98% and an f1-score of 0.98 or 98%.

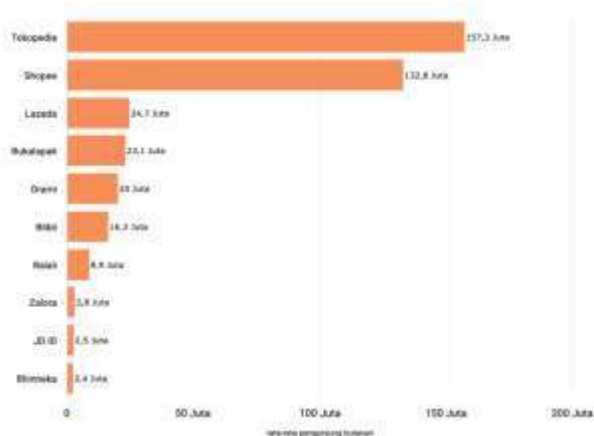
Keyword— SVM, Shopee, Sentiment Analysis

I. PENDAHULUAN

Indonesia merupakan salah satu negara dengan pertumbuhan *e-commerce* yang sangat pesat. Sebagian besar masyarakatnya pernah melakukan transaksi belanja secara *online*. Banyaknya transaksi *online* tersebut dilakukan melalui berbagai macam platform *marketplace* seperti Lazada, Tokopedia, Shopee, Bukalapak, dan berbagai macam *marketplace* lainnya. Tren *marketplace* terus berkembang seiring perkembangan internet dan teknologi digital di Indonesia. Minat dan antusiasme masyarakat terhadap perkembangan ini menyebabkan semakin banyak jumlah *marketplace* yang bermunculan di Indonesia. Salah satu *marketplace* yang saat ini banyak diminati di Indonesia adalah Shopee. Berdasarkan data yang diperoleh dari databoks, Shopee merupakan *marketplace* yang berada di peringkat kedua setelah Tokopedia [1].

Kemajuan teknologi menyebabkan toko online dapat diakses melalui website atau aplikasi, sehingga saat ini kita hanya perlu berbelanja melalui *gadget* tanpa keluar rumah. Selain hemat waktu dan tenaga, banyaknya promosi dan harga yang lebih murah dibandingkan membeli di toko biasa membuat orang senang berbelanja *online* [2].

Banyak start-up di bidang *e-commerce* berlomba-lomba membuat aplikasi toko online. Berikut urutan 10 *e-commerce* berdasarkan jumlah pengunjung terbanyak di Indonesia kuartal I 2022 [3]:



Gambar 1. Pengunjung E-Commerce Kuartal I 2022

Shopee menjadi toko online dengan pengunjung terbanyak ke dua setelah Tokopedia, yaitu 132,8 juta pengunjung pada kuartal I 2022, mengalahkan beberapa toko online terkenal lainnya. Jumlah kunjungan web Shopee di Indonesia meningkat +117% dibandingkan jumlah kunjungan pada kuartal pertama tahun 2018. Perusahaan yang berbasis di Singapura ini mencatat rata-rata 74.995.300 kunjungan per bulan ke situs pertamanya selama kuartal tahun 2019. Pada kuartal pertama 2018, Shopee hanya mampu mengimport sekitar 34.510.800 pengunjung di situs versi Indonesia. Peningkatan rata-rata pengunjung web juga meningkatkan posisi Shopee dari tahun ke tahun, dari posisi 8 di Q1 2017 menjadi posisi 3 di peta elektronik Q1 2019[4] dan di tahun 2022 berada di posisi 2. Namun, aplikasi toko online tentu memiliki kelebihan dan kekurangan, termasuk Shopee. Pengalaman pengguna Shopee sebagai aplikasi dengan pengguna paling aktif kedua sering diungkapkan di kolom komentar di *Play Store* baik berupa kritik maupun kepuasan. Perasaan negatif atau positif pengguna tentang suatu produk berhubungan dengan berbagai aspek produk. Misalnya, aplikasi Shopee, meskipun tidak terlihat bagus, tetap populer karena fiturnya yang mudah digunakan.

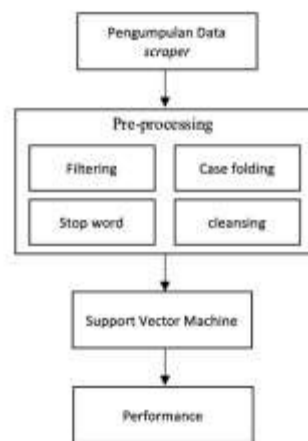
Masalah ini disebut *Aspect-Based Sentiment Analysis* (ABSA). Penelitian sebelumnya telah dilakukan pada data review film[5]. Teks ulasan tidak hanya menjelaskan nuansa umum film, tetapi menunjukkan berbagai aspek seperti suara, lirik, kualitas rekaman, kreativitas, dan lain-lain. Contoh kalimatnya adalah “Saya suka ceritanya, tapi tidak suka musiknya” kalimat tersebut mengandung dua perasaan yang berlawanan dari dua aspek, yaitu cerita dan musik. Penelitian ini melakukan pemeringkatan sentimen dengan mengklasifikasikan setiap kata sebagai positif atau negatif, dengan mengambil skor rata-rata semua kata di SentWordNet. Jika rata-rata lebih besar dari 0, kalimat dinilai sebagai sentimen positif, sebaliknya negatif. Proses klasifikasi aspek menggunakan tag *semantic* yang dianotasi dengan proses *semantic*. Pencarian lainnya juga dilakukan pada data blog berbahasa Mandarin[6], pencarian tema global dilakukan menggunakan model *Latent Dirichlet Allocation* (LDA) kemudian untuk tema yang lebih spesifik yaitu tema *local* menggunakan metode *sliding window*. Penelitian ini menunjukkan akurasi sebesar 92,15%. Penelitian lainnya tentang Analisis Sentimen Terhadap Acara Televisi Berdasarkan Opini Publik. Penelitian ini melakukan

klasifikasi terhadap acara televisi dan menghasilkan opini yang dibagi ke dalam dua kelas positif dan negatif, dengan menggunakan Algoritma *Naive Bayes Classifier* serta pengujian menggunakan algoritma *percentage split* dan diperoleh hasil akurasi sebesar 90% [7].

Berdasarkan permasalahan diatas, penelitian ini akan dilakukan dengan bantuan Algoritma *Support Vector Machine* (SVM) sebagai algoritma pengklasifikasi yang mampu melakukan klasifikasi terhadap komentar pengguna aplikasi Shopee ke dalam komentar negatif dan positif. Algoritma *Support Vector Machine* merupakan algoritma klasifikasi yang lebih unggul dibanding Metode *K-Nearest Neighbor* (KNN) dengan perolehan akurasi mencapai 89,70% [8]. Bahkan pada penelitian yang dilakukan Primandani Arsi dan Retno Waluyo SVM mampu menghasilkan akurasi sebesar 96,68% [9].

II. METODE

Penelitian ini menggunakan penelitian deskriptif, dengan melakukan pemecahan masalah terhadap kondisi pokok ataupun pada penelitian berdasarkan fakta yang ada. Berikut tahapan penelitian yang ada pada gambar 2 :



Gambar 2. Tahapan Penelitian

A. Text Mining

Text Mining merupakan proses mengekstraksi suatu pola untuk diteliti yang datanya berasal dari sebuah teks. *Text Mining* merupakan disiplin ilmu yang didasarkan pada *information retrieval*, *data mining*, *machine learning*, *statistic* dan *linguistic* komputasi[10].

B. Preprocessing

Tahapan *preprocessing* yang dilakukan dalam penelitian ini :

1. Mengubah kata Slang

Slang merupakan kata-kata tidak baku yang bersifat musiman dan biasanya digunakan oleh remaja atau kelompok *social* tertentu sehingga hanya kelompok tersebut yang dapat memahami komunikasinya[11]. Kata-kata di kolom komentar seringkali mengandung jargon atau akronim, sehingga jargon atau akronim perlu diganti dengan kata-kata yang mudah kita pahami. Untuk mengubah kata slang menjadi kata baku diperlukan sebuah alat bantu berupa kamus kata.

2. Case-folding

Case-folding mengubah semua karakter menjadi sama, misalnya kata-kata diubah dari huruf besar ke

huruf kecil[10]. Untuk mengubah semua data ke huruf kecil menggunakan fungsi *string.lower()*,

sedangkan untuk menggubah data ke huruf besar menggunakan fungsi *string.upper()* atau bisa juga membuat fungsi sendiri untuk mengonversi semua huruf menjadi besar atau kecil. [13]

3. Menghapus *Stopword*

Stopword merupakan kata yang tidak memiliki semantic dan tidak merujuk pada informasi yang relevan dengan kasus yang diteliti, umumnya kata depan dan topik [10]. Perlu dibuat *directory* khusus untuk kata-kata *stopword*, yang dapat diperbarui sesuai dengan data yang akan diperiksa. Contoh *stopword* dalam Bahasa Inggris seperti “the”, “a”, “when”, “is”, dan lain-lain.

4. Menghapus Simbol (*cleansing*)

Karena kumpulan komentar apa pun, baik di web atau di jejaring *social*, sering kali berisi karakter non-huruf, komentar tersebut harus dihapus karena pemrosesannya juga tidak penting untuk analisis data.

C. Support Vector Machine (SVM)

Support Vector Machine (SVM) adalah Teknik pembelajaran dengan banyak kualitas yang diinginkan dan menjadikan algoritma SVM sangat populer. SVM mempunyai dasar teoritis yang kuat dan melakukan klasifikasi lebih akurat daripada kebanyakan algoritma lain di banyak aplikasi. Banyak penelitian telah melaporkan bahwa SVM merupakan metode yang paling akurat untuk klasifikasi teks[12]. SVM juga banyak digunakan dalam klasifikasi sentimen[5].

III. HASIL DAN PEMBAHASAN

A. Data Ulasan

Total data ulasan yang diperoleh melalui proses *scraping*, terdapat 3000 data ulasan terbaru. Berikut contoh data yang berhasil di angkat menggunakan metode *scraping*:

TABEL 1. HASIL SCRAPING

review Id	User Name	content	score	at	Reply Content
a8b33749-c113-41d4-8d35-4c05978c45b0	Said Ajih	Biarkan bintang bicara.	5	22/10/22 13.18	Hai kak , makasih ya buat review positifnya. M...
1be7c6e9-7407-4d6e-9821-1b1462c3c191	odik simanto	kembalikan shopee ke versi lama , apk mulai ga...	1	22/10/22 13.18	Hi kak, mohon maaf buat km ganyaman. Terkait a...
a5336f16-f8e9-42db-8e82-67ffc972df12	Jayanti Ida	CUSTOMER CARE BURUK SEKALI PENYELESAIAN MASALA...	1	22/10/22 13.17	hi kak, maaf ya atas kendalanya. Shopee terus ...

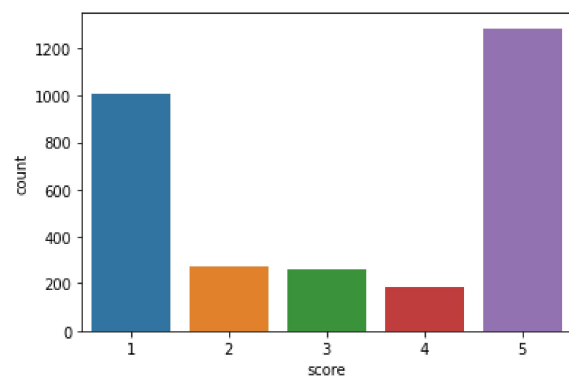
5a1de657-c2a4-404c-9782-7de5b5a8cfbd	Superkopi Roastery	Jolke tahu wae, Lik . .	3	22/10/22 13.17	hii kak, maaf ya atas kendalanya. Shopee terus...
f94b20ad-f259-483a-b986-3215972b2f58	Neni Nine	Kok lelet banget sih kirain karena sinyal tern...	1	22/10/22 13.17	Halo kak , maaf ya udah buat kamu ga nyaman. M...

Untuk melakukan pelabelan, pada penelitian ini akan melihat dari *rating* pengguna melalui ulasan pada aplikasi Shopee. Proses pelabelan akan di bagi menjadi dua yaitu label positif dilihat dari pemberian *score* di atas 3 sampai 5 sedangkan untuk label negatif dari *score* 1 sampai dengan 3. Sebagai contoh pada tabel 1 diatas *content* “Kok lelet banget sih kirain karena sinyal” dengan *score* 1 maka akan di label kan negatif, sedangkan *content* “Biarkan bintang bicara” dengan *score* 5 akan dilabelkan positif.

TABEL 2. JUMLAH ULASAN TIAP RATING

Rating	Jumlah
1	1005
2	270
3	258
4	185
5	1282
Total	3000

Berdasarkan tabel 2. Jumlah ulasan tiap *rating*, rating 5 mempunyai jumlah terbanyak dari *rating* lainnya yaitu 1282 ulasan disusul dengan rating 1 sebanyak 1005 ulasan dan rating 4 memiliki jumlah terendah yaitu 185 ulasan.



Gambar 3. Grafik Rating Ulasan

Setelah pelabelan selesai, langkah selanjutnya menghilangkan variabel yang tidak diperlukan sehingga akan menyisakan variabel yang dibutuhkan saja yaitu *content* dan sentimen, berikut contoh data setelah dilakukan *drop* variabel yang tidak di butuhkan :

TABEL 3. CONTOH DATA SETELAH DILAKUKAN DROP VARIABEL

No	content	sentimen
0	Biarkan bintang bicara.	Positif
1	kembalikan shopee ke versi lama , apk mulai ga...	Negatif
2	CUSTOMER CARE BURUK SEKALI PENYELESAIAN MASALA...	Negatif
3	Jolke tahu wae, Lik . .	Negatif
4	Kok lelet banget sih kirain karena sinyal term...	Negatif

B. Pre-Processing

Pada proses ini dilakukan *filtering*, *case-folding*, *stopword* dan *cleansing* atau pembersihan. Berikut data yang telah dilakukan proses *pre-processing* :

TABEL 4. DATA SETELAH PROSES PRE-PROCESSING

No	Content	Sentimen
1250	alhamdulillah	Positif
1927	aplikasi belanja online pale suka lemoooot kadan...	Positif
2560	mantep	Positif
1783		Positif
1048	Bagu	Positif
...
2221	tolong update jaringan lemot padah buka aplikas...	Negatif
1086	lemot aplikasi nya jaringan dah kenceng coba c...	Negatif
995	shopee rudet hoream	Negatif
63	padah sinyal bagus aplikasi lancar yg load tru...	Negatif
235	updat wara update tolol	Negatif

Untuk mengetahui kata-kata yang sering muncul atau visualisasi data di label negatif dan positif, digunakan fungsi *wordcloud*. Tampilannya seperti dibawah, *wordcloud* warna biru label positif sedangkan *wordcloud* warna merah label negatif :



Gambar 4. Wordcloud label Positif dan Negatif

Langkah selanjutnya dilakukan proses SVM dengan menggunakan parameter $C = 0.01, 0.05, 0.25, 0.5, 0.75, 1$ sehingga di dapat nilai tertinggi :

Akurasi untuk $C = 0.01$: 0.9063032367972743

Akurasi untuk $C = 0.05$: 0.9557069846678024

Akurasi untuk $C = 0.25$: 0.979557069846678

Akurasi untuk $C = 0.5$: 0.9812606473594548

Akurasi untuk $C = 0.75$: 0.9812606473594548

Akurasi untuk $C = 1$: 0.9812606473594548

Dari nilai diatas didapatkan nilai terbaik dari nilai C yang diberikan adalah $C = 0.75$: 0.9812606473594548 akurasi yang didapatkan adalah 98 %

TABEL 5. HASIL EVALUASI KINERJA SVM

	precision	recall	f1-score	support
Negatif	0.99	0.97	0.98	285
Positif	0.97	0.99	0.98	302
Accuracy			0.98	587
Macro avg	0.98	0.98	0.98	587
Weighted avg	0.98	0.98	0.98	587

IV. KESIMPULAN

Setelah dilakukan penelitian dengan menggunakan model algoritma *Support Vector Machine* (SVM) dapat disimpulkan bahwa, SVM terbukti sebagai algoritma klasifikasi yang cukup baik untuk menganalisis sentimen pada pengguna aplikasi shopee dengan perolehan akurasi sebesar 98% dan f1-score 0.98 atau sebesar 98%.

REFERENSI

- [1] C. Cahyaningtyas, Y. Nataliani, I. R. Widiarsari, F. T. Informasi, U. Kristen, and S. Wacana, "Analisis sentimen pada rating aplikasi Shopee menggunakan metode Decision Tree berbasis SMOTE," vol. 18, no. 2, pp. 173–184, 2021.
- [2] C. A. Sukandar, "Ternyata Ini 5 Alasan Millenial Gemar Belanja Online, Enggak Heran Deh!," 2018. <https://wartaekonomi.co.id/read203120/ternyata-ini-5-alasan->.
- [3] V. A. Dihni, "10 E-Commerce dengan Pengunjung Terbanyak Kuartal I 2022," 2022. <https://databoks.katadata.co.id/datapublish/2022/07/19/10-e-commerce-dengan-pengunjung-terbanyak-kuartal-i-2022> (accessed Oct. 13, 2022).
- [4] A. Luthfi, "11 Temuan Penting Peta E-Commerce Indonesia di Q1 2019," 2019.
- [5] T. T. Thet, J. Na, and C. S. G. Khoo, "Aspect-based sentiment analysis of movie reviews on discussion boards," vol. 36, no. 6, pp. 823–848, 2010, doi: 10.1177/0165551510388123.
- [6] F. Xianghua, L. Guo, G. Yanyan, and W. Zhiqiang, "Knowledge-Based Systems Multi-aspect sentiment analysis for Chinese online social reviews based on topic modeling and HowNet lexicon," *Knowledge-Based Syst.*, vol. 37, pp. 186–195, 2013, doi: 10.1016/j.knosys.2012.08.003.
- [7] Sentiaji, A. R., & Bachtiar, A. M. (2014). Analisis sentimen terhadap acara televisi berdasarkan opini publik. *Jurnal Ilmiah Komputer dan Informatika (KOMPUTA)*, 2(1), 55-60.
- [8] Nasution, M. R. A., & Hayaty, M. (2019). Perbandingan Akurasi dan Waktu Proses Algoritma K-NN dan SVM dalam Analisis Sentimen Twitter. *Jurnal Informatika*, 6(2), 226-235.
- [9] Arsi, P., & Waluyo, R. (2021). Analisis Sentimen Wacana Pemindahan Ibu Kota Indonesia Menggunakan Algoritma Support Vector Machine (SVM). *Jurnal Teknologi Informasi Dan Ilmu Komputer*, 8(1), 147.
- [10] I. S. K. Idris and Y. A. Mustofa, "Typo Checking Menggunakan Algoritma Rabin-Karp," *Jambura J. Electr.*, 2022, [Online]. Available: <https://ejurnal.ung.ac.id/index.php/jjee/article/view/12150>.
- [11] Kamus versi online/daring (dalam jaringan), "Kamus Besar Bahasa Indonesia (KBBI)," 2022. <https://kbbi.web.id/slang> (accessed Oct. 23, 2022).
- [12] T. N. Wijaya, R. Indriati, and M. N. Muzaki, "Analisis Sentimen Opini Publik Tentang Undang-Undang Cipta Kerja Pada Twitter," *Jambura J. Electr. Electron. Eng.*, vol. 3, no. 2, pp. 78–83, 2021, doi: 10.37905/jjee.v3i2.10885.
- [13] B. Liu, *Web Data Mining*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011.
- [14] M. Tsytsarau and T. Palpanas, "Survey on mining subjective data on the web," *Data Min. Knowl. Discov.*, vol. 24, no. 3, pp. 478–514, 2012, doi: 10.1007/s10618-011-0238-6.