

ASPECT EXTRACTION OF E-COMMERCE AND MARKETPLACE APPLICATIONS USING WORD2VEC AND WORDNET PATH

Dhani Ratna Sari^{*1}, Bayun Matsaany², Muhammad Hamka³

¹Digital Business, Institut Teknologi dan Bisnis Muhammadiyah Purbalingga, Indonesia

²Actuarial, Institut Teknologi dan Bisnis Muhammadiyah Purbalingga, Indonesia

³Informatics Engineering, Fakultas Teknik dan Sains, Universitas Muhammadiyah Purwokerto, Indonesia

Email: ¹dhani_ratnasari@itbmp.ac.id, ²bayunmatsaany@itbmp.ac.id, ³muhammadhamka@ump.ac.id

(Article Received: Desember 7, 2022; Revision: January 1, 2023; Published: August 18, 2023)

Abstract

Aspect extraction is an essential element in Aspect-Based Sentiment Analysis (ABSA). Errors in determining aspects of ABSA will result in errors in determining the sentiment of an opinion and the accuracy value of ABSA. This study aims to obtain elements of opinion sentences on using e-commerce applications and marketplaces in Indonesia. Corrections of the statement were sourced from social media Twitter with the keywords "e-commerce" and "marketplace" from August 2020 to January 2022, and a total of 54,244 comments were obtained. Determination of the words that are candidate aspects is selected using POS Tagging for classes of noun singular (NN), noun plural (NNS), proper noun singular (NNP), and proper noun plural (NNPS). Furthermore, the selected word is given weight using TF-IDF. Calculating similarity values between words is obtained using cosine similarity Word2Vec with CBOW architecture. Next, the Word2Vec results are compared with the closeness of meaning obtained from WordNet's path_similarity gold standard, and the correlation is calculated using the Pearson Correlation equation. The feature extraction results obtained 14 aspects in Word2Vec with windows size 6 and vector dimension 150. The highest word similarity correlation between Word2Vec and WordNet gold standard was achieved in Word2Vec using windows size 9 and vector dimension 300 with a value of 0.596.

Keywords: cosine similarity, aspect extraction, gold standard wordnet, natural language processing, pos tagging.

ASPECT EXTRACTION PENGGUNAAN APLIKASI E-COMMERCE DAN MARKETPLACE MENGGUNAKAN WORD2VEC DAN PATH WORDNET

Abstrak

Ekstraksi aspek merupakan elemen penting di dalam Aspect-Based Sentiment Analysis (ABSA). Kesalahan penentuan aspek pada ABSA akan berakibat pada kesalahan penentuan sentimen suatu opini serta nilai akurasi ABSA. Penelitian ini bertujuan mendapatkan aspek-aspek dari kalimat opini penggunaan aplikasi e-commerce dan marketplace di Indonesia. Kalimat opini bersumber dari media sosial Twitter dengan kata kunci "e-commerce" dan "marketplace" sejak bulan Agustus 2020 hingga Januari 2022 dan diperoleh sebanyak 54.244 kalimat ulasan. Penentuan kata-kata yang menjadi calon aspek dipilih menggunakan POS Tagging untuk kelas kata noun singular (NN), noun plural (NNS), proper noun singular (NNP) dan proper noun plural (NNPS). Selanjutnya kata yang terpilih diberikan bobot menggunakan TF-IDF. Perhitungan nilai kemiripan antar kata diperoleh dengan mengimplementasikan cosine similarity Word2Vec dengan arsitektur CBOW. Selanjutnya, hasil Word2Vec dibandingkan dengan kedekatan makna yang diperoleh dari path_similarity gold standard WordNet dan dihitung korelasinya menggunakan persamaan Pearson Correlation. Hasil ekstraksi fitur diperoleh 14 aspek pada Word2Vec dengan ukuran windows size 6 dan vector dimension 150. Korelasi kemiripan kata hasil Word2Vec dan gold standard WordNet tertinggi dicapai pada Word2Vec menggunakan parameter windows size 9 dan vector dimension 300 dengan nilai 0.596.

Kata kunci: cosine similarity, ekstraksi aspek, gold standard wordnet, natural language processing, pos tagging.

1. PENDAHULUAN

E-Commerce telah menjadi salah satu pasar menjanjikan untuk transaksi penjualan barang dan jasa selama dua dekade terakhir ini. Sementara itu efek pandemi COVID-19 telah mengubah cara masyarakat melakukan transaksi belanja. Laporan survei dari We Are Social menyatakan terjadi peningkatan jumlah penduduk Indonesia yang melakukan transaksi belanja *online* sebesar 14,9% atau 21 juta penduduk [1]. Lebih lanjut Hasil survei Badan Pusat Statistik menyebutkan terjadi peningkatan sekitar 5,76% perusahaan baru yang memanfaatkan internet dan Teknologi Informasi untuk pemasaran, dan 4 (empat) dari 5 (lima) pelaku usaha atau sekitar 80 % menyatakan bahwa penggunaan internet dan TI serta pemasaran secara *online* berdampak signifikan terhadap tingkat penjualan produk mereka [2].

Peningkatan penggunaan *e-commerce* tersebut diikuti penambahan data teks yang bersifat subjektif dalam jumlah besar dan berisi sentimen dan opini pengguna *e-commerce* melalui internet [3], khususnya di media sosial. Media sosial dalam satu dekade terakhir telah menjadi *platform* komunikasi digital di mayoritas organisasi [4]. Melalui media sosial pengguna dapat secara mudah membagikan informasi, berinteraksi dengan pengguna lain yang memiliki minat dan tujuan yang sama, atau memberikan opini atau review tentang suatu produk atau layanan. Peran media sosial saat ini tidak hanya menghubungkan antar pengguna melalui jaringan internet, akan tetapi membantu perusahaan, organisasi, atau pelaku bisnis mengembangkan hubungan yang lebih dengan konsumen atau kliennya [5]. Lebih lanjut survei yang dilakukan oleh Wham (2018) dalam [5] menyebutkan bahwa 76% konsumen memilih pembelian produk atau layanan yang ditawarkan melalui media sosial.

Beberapa tahun terakhir, peran sosial media menjadi penting. Pada Juni 2022, konten internet yang sering diakses di Indonesia adalah media sosial dengan persentase mencapai 89,15% [6] dan jumlah pengguna mencapai 191,4 juta [1]. Jumlah konten di media sosial yang begitu besar memberikan peluang penggunaannya sebagai sumber data untuk berbagai kajian penelitian. Lebih lanjut, studi terkait analisis media sosial memberikan pemahaman mendalam tentang fenomena sosial yang terjadi dan dapat digunakan untuk menghasilkan berbagai peluang bisnis, seperti keunggulan kompetitif [7].

Meskipun analisis data media sosial dapat menjadi sumber pengambilan keputusan berbasis data, akan tetapi dengan besarnya volume data dari media sosial, proses untuk mendapatkan informasi yang tepat menimbulkan kendala dalam pengelolaan data, karena membutuhkan banyak waktu dan keahlian khusus. Sehingga dibutuhkan pendekatan *Natural Language Processing* (NLP) untuk memudahkan proses ekstraksi informasi atau pengetahuan dari teks opini di media sosial. Salah

satunya bagian dari NLP yang digunakan untuk menghasilkan klasifikasi informasi berdasarkan teks review adalah *Aspect Based Sentiment Analysis* (ABSA).

ABSA tidak hanya mengklasifikasikan polaritas sentimen teks opini, tetapi juga mengekstraksi aspek atau *term* [8]. Sehingga penentuan polaritas sentimen pada ABSA sangat bergantung berdasarkan aspek yang didapatkan dari kalimat ulasan pengguna. Dalam ABSA komponen *Aspect Term Extraction* (ATE) merupakan langkah pertama dan utama karena terkait dengan hasil akurasi dan penilaian sentimen berdasarkan aspek di dalam ABSA [9]. Ekstraksi aspek memiliki pengaruh besar pada akurasi ABSA karena kesalahan pada saat identifikasi aspek atau *term* akan berpengaruh pada langkah berikutnya di dalam ABSA dan menyebabkan penentuan sentimen menjadi tidak tepat. Tujuan ATE adalah melakukan transformasi dari sumber teks tidak terstruktur menjadi representasi informasi terstruktur. Dalam prosesnya dilakukan proses penghapusan fitur yang tidak berkaitan dengan suatu topik dan fitur yang terdapat duplikasi [10], berikutnya fitur-fitur penting akan dikombinasikan dengan masing-masing bobotnya sebagai representasi informasi yang terkandung dalam teks ulasan.

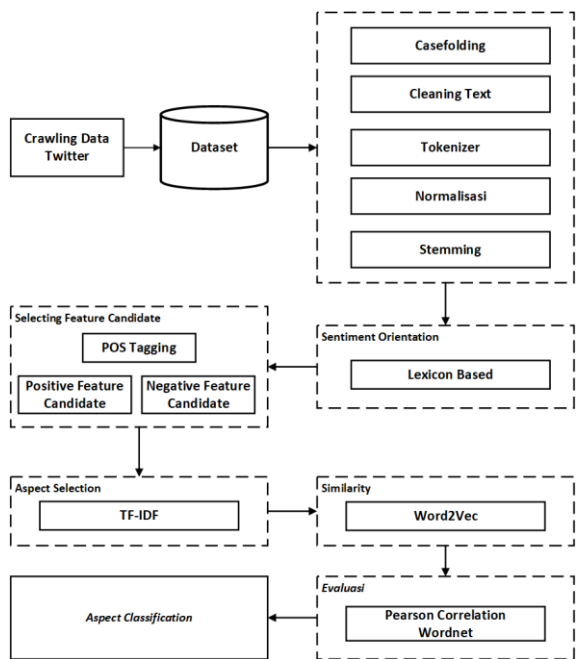
Metode representasi teks yang sering digunakan untuk menemukan informasi pada teks adalah BoW (*bag of words*) dan TF-IDF (*term frequency-inverse document frequency*) [10]. Pada model BoW himpunan kata-kata yang sudah ditentukan disebut sebagai "*dictionary*", kemudian setiap dokumen diwakili oleh vektor histogram yang menghitung jumlah kemunculan masing-masing kata dalam dokumen [11]. Vektor tiap kata di dalam dokumen direpresentasikan sebagai $v = [x_1, x_2, \dots, x_n]$, dimana x_i merupakan jumlah kemunculan kata ke- i pada kata atau *term* yang sudah ditentukan [12]. Representasi vektor kata pada makalah ini menggunakan mekanisme pembobotan TF-IDF [13], TF-IDF juga digunakan untuk ekstraksi fitur. Ekstraksi fitur atau aspek yang dihasilkan dimungkinkan memiliki kesamaan makna, sehingga diperlukan pemanfaatan proses *similarity word* untuk menentukan fitur yang memiliki makna yang sama. Teknik penentuan kesamaan makna fitur menggunakan *word embeddings Word2Vec*. *Word2Vec* memiliki keunggulan terutama pada model yang mempertimbangkan urutan kata dalam dokumen dibandingkan model BoW atau *N-Gram Bag* [14].

Penelitian ini mengimplementasikan TF-IDF untuk perhitungan bobot kandidat *term* atau aspek untuk ABSA penggunaan aplikasi *e-commerce* dan *marketplace*. Kata yang dipilih menjadi kandidat fitur adalah kata dengan label kata benda atau *noun singular* (NN), *noun plural* (NNS), *proper noun singular* (NNP) dan *proper noun plural* (NNPS). Kemudian berdasarkan kandidat fitur, dipilih 50 kata

dengan bobot TF-IDF tertinggi pada klasifikasi sentimen positif dan negatif sebagai aspek. Kemudian tiap-tiap aspek pada sentimen positif dan negatif akan dihitung kemiripannya berdasarkan nilai *cosine distance* menggunakan *Word2Vec* dengan arsitektur *Skip-Gram*. Semakin tinggi nilai *cosine distance* menunjukkan orientasi semantik kata yang lebih dekat [14]. Melalui arsitektur *Skip-Gram*, suatu *term* akan diprediksi untuk mendapatkan kata-kata yang memiliki kedekatan ruang vektor. Kata yang memiliki kedekatan makna akan berada di ruang vektor yang sama, sedangkan kata yang tidak memiliki kemiripan akan berada pada ruang vektor yang berbeda [15].

2. METODE PENELITIAN

Ekstraksi aspek atau *term* dari kalimat ulasan di media sosial adalah tugas yang kompleks, karena melibatkan penanganan dan analisis data yang tidak terstruktur dalam jumlah besar [16]. Proses yang dilakukan pada ekstraksi aspek meliputi 1) *Crawling Twitter Data*, 2) *preprocessing*, 3) *vectorization*, 4) *sentiment orientation*, 5) *selecting feature candidate*, 6) *aspect selection*, dan 7) *similarity*. Tahapan penelitian ditunjukkan pada gambar 1.



Gambar 1. Tahap Penelitian

Data yang digunakan dikumpulkan melalui *crawling* media sosial Twitter berisi kalimat ulasan dengan kata kunci "e-commerce" dan "marketplace". Kalimat ulasan dikumpulkan dari bulan Agustus 2020 hingga Januari 2022. Sebelum dianalisis, data ulasan tersebut dilakukan *preprocessing* untuk membersihkan data dari karakter tidak valid [17], mengubah kata menjadi bentuk baku, serta menghilangkan kata yang tidak memiliki pengaruh terhadap analisis sentimen.

Proses *preprocessing* terdiri dari beberapa tahap, yaitu *cleaning text*, *casefolding* untuk mengubah semua karakter menjadi huruf kecil, *tokenizing* untuk memisahkan tiap kata dari kalimat ulasan, *normalisasi* untuk mengubah kata menjadi bentuk baku, *filtering* untuk menghilangkan tanda baca, hashtag, URL, identitas akun, simbol, dan lain sebagainya, *stemming* untuk mengubah kata dalam bentuk dasarnya dan *stopword removing* untuk menghapus kata-kata yang tidak memiliki keterkaitan dalam analisis sentimen [18].

Tahap selanjutnya adalah memberikan label sentimen pada tiap kalimat ulasan. Pendekatan yang dilakukan adalah *lexicon* dengan kamus *Senticnet 7* [19]. Pelabelan sentimen dilakukan dengan membuat *function* untuk menghitung nilai polaritas suatu kalimat opini. Penentuan nilai polaritas suatu ulasan, digunakan persamaan (1) [20]. Berdasarkan *score* opini, ditentukan polaritas kalimat apakah positif, negatif atau netral melalui persamaan (2) [21].

$$Score = \sum_{i=1}^n Score + W_{positive} + W_{negative} \tag{1}$$

$$Score \begin{cases} \text{Positif, jika } Score > 0 \\ \text{Netral, jika } Score = 0 \\ \text{Negatif, jika } Score < 0 \end{cases} \tag{2}$$

Proses pemilihan *feature candidate* dikelompokkan sesuai *dataset* yang sudah diberi label sentimen. Kata-kata yang menjadi kandidat fitur adalah kata dengan kelas kata benda atau *noun singular (NN)*, *noun plural (NNS)*, *proper noun singular (NNP)* dan *proper noun plural (NNPS)*. Sesuai kandidat aspek yang berhasil diidentifikasi, tahap berikutnya adalah pemilihan aspek. Pemilihan aspek dilakukan berdasarkan hasil pembobotan tiap kata pada kandidat fitur menggunakan *Term Frequency (TF)* dan *Inverse Document Frequency (IDF)*. Penentuan nilai bobot tiap kata dalam kandidat aspek dihitung menggunakan persamaan (3), (4), dan (5) [22].

$$TF - IDF(d, t) = TD(d, t) * IDF(t) \tag{3}$$

$$TF(d, t) = \frac{\text{jumlah kata } t \text{ pada dokumen } d}{\text{total kata pada dokumen } d} \tag{4}$$

$$IDF(t) = \log \frac{\text{total dokumen}}{\text{dokumen mengandung kata } t} \tag{5}$$

Berdasarkan nilai bobot tiap kata, ditentukan masing-masing 50 aspek pada sentimen positif dan negatif. Tiap aspek kemudian dilakukan proses perhitungan kemiripan secara semantik (*semantic similarity*) menggunakan *Word2Vec*. *Semantic similarity* adalah konsep yang dapat mengukur kesamaan makna sesuai kotneks dalam suatu teks [23]. *Word2Vec* bekerja dengan mengubah kata dalam bentuk representasi vektor. Selanjutnya nilai kemiripan antar kata dibangkitkan menggunakan persamaan *Cosine Similarity* dari nilai vektor kata yang dihasilkan oleh model *Word2Vec* [24]. Perhitungan *Cosine Similarity* ditunjukkan pada

persamaan (6). Arsitektur Word2Vec yang digunakan dalam penelitian ini adalah *Skip-Gram* dengan data latih bersumber dari korpus Wikipedia Bahasa Indonesia dengan jumlah artikel sebanyak 469.623.

$$similarity = \cos \theta = \frac{\bar{x} \bullet \bar{y}}{\|\bar{x}\| \|\bar{y}\|} \quad (6)$$

Dimana :

$x \bullet y$: vector dot product dari x dan y, dihitung dengan $\sum_{k=1}^n x_k y_k$

$\|x\|$: panjang vektor x, dihitung dengan $\sum_{k=1}^n x_k^2$

$\|y\|$: panjang vektor x, dihitung dengan $\sum_{k=1}^n y_k^2$

Hasil perhitungan kemiripan antar term dievaluasi menggunakan *Pearson Correlation* terhadap hasil *path_similarity gold standard* Wordnet. *Pearson Correlation* dihitung menggunakan persamaan (7) [23]

$$r = \frac{n \sum xy - (\sum x)(\sum y)}{\sqrt{n \sum x^2 - (\sum x)^2} \sqrt{n \sum y^2 - (\sum y)^2}} \quad (7)$$

Dimana :

- n : jumlah pasangan kata
- x : nilai kemiripan dari sistem (Word2Vec)
- y : nilai kemiripan berdasarkan *gold standard*

Sedangkan kriteria korelasi kedekatan antar kata ditentukan berdasarkan 0 hingga 1. Klasifikasi korelasi kedekatan antar kata ditunjukkan pada tabel 1 [25].

Tabel 1. Kriteria *Pearson Correlation*

Nilai r	Kriteria Korelasi
0	Tidak ada korelasi
0-0,5	Korelasi lemah
0,5-0,8	Korelasi sedang
0,8-1	Korelasi kuat
1	Korelasi sempurna (dua kata sama)

3. HASIL DAN PEMBAHASAN

Pengumpulan data ulasan *Twitter* terkait *e-commerce* dan *marketplace* sepanjang Agustus 2020 hingga Januari 2022 berjumlah 54.244 ulasan. Fitur data meliputi *id*, *created_at*, dan *text*. Data yang dikumpulkan merupakan data tidak terstruktur, karena berbentuk kalimat opini bebas. Tampilan potongan data ditunjukkan pada tabel 2.

Tabel 2. Data Opini *Twitter*

No	Text	Created_at
0	POLOSE juga bisa kalian dapatkan melalui akun E-Commerce Bukalapak, Tokopedia, dan Shopee #kaospolos #polosejogja @akuyakuu_ @airyrabbit	2020-08-09T23:43:01.000Z
1	@txtdarionlshop Yups bener soalnya di apk e-commerce sebelah udh bisa ðŸ™, @zakialir @gitch3	2020-08-09T20:42:39.000Z
2		2020-08-

	@collegemenfess Iniii, beli di e-commerce lebih murah https://t.co/OS8H54pw54 Gue nemu "himalaya neem purifying mask" ini di tiptop depok 2, harganya juga lebih murah dr e-commerce juga mell #fyibuayanglainjuga https://t.co/zq0BxBOAyZ	09T18:23:50.000Z
3	Hari Belanja Diskon pacu perkembangan bisnis UMKM bersama æœe-commerceæœe https://t.co/v8UY8CVwNI	2020-08-09T18:04:29.000Z
4
...
54239	Tau gak guys?? Kalo KRASmart di buat seperti marketplace retail pada umumnya, agar memudahkan para pembeli #BeliBajadiKRASmart https://t.co/2eL5xZkvoq	2021-11-27T02:52:44.000Z
54240	LOKER JOGJA TERBARU: Content Writer - Admin Marketplace di PT. Gawe Becik Nadhah Anugrah (GENAH) #lokerjogja #lowongankerjajogja #loker #jogja #yogyakartahttps://t.co/934F16Jq5Q	2021-11-27T02:50:54.000Z
54241	Nah guys kalian udah pada tau ga sih kalau KRASmart ini di buat ya sama seperti marketplace retail pada umumnya biar memudahkan para pembeli #BeliBajadiKRASmart https://t.co/1LeawwpKTT	2021-11-27T02:42:46.000Z
54242	Jangan lewatkan Pameran Buku Hybrid Edition Indonesia International Book Fair 2021 'IT'S A BOOK AFFAIR' ! 08 Desember-12 Desember 2021 Offline : Cendrawasih Hall Senayan JCC 10.00 WIB s.d 20.00 WIB Online : IIBF 2021 Shopee Official Marketplace : shopee_id 24 JAM NONSTOP https://t.co/hhy6wIjVZ8	2021-11-27T02:30:08.000Z
54243	ðŸ† Malam ni Jen Mai marketplace pentas ðŸ~ ðŸœ~ Alhamdulillah, banyak dah siap ðŸ†ðŸ™œðŸ• »ã• seronoknya! After struggle near a month, finally~ ðŸ~ðŸ~ðŸ~ ðŸ• ðŸ• » Stay tuned naaa ðŸ†ðŸœ ðŸœ• ðŸœ¼ã•ãœ https://t.co/zZWOPePzZq	2021-11-27T02:15:29.000Z

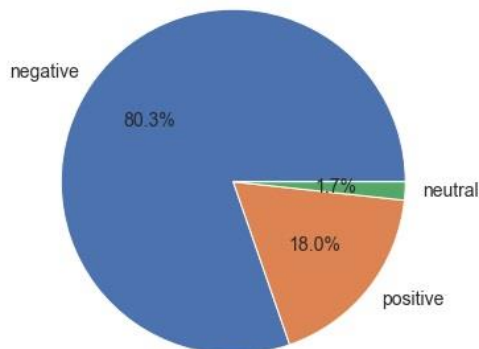
Data pada tabel 2 merupakan data yang masih terdapat *noise*, sehingga perlu dilakukan *preprocessing* untuk membersihkan data. Pembersihan data juga melibatkan penghapusan duplikasi ulasan serta menghilangkan kata yang hanya terdiri dari 1 karakter, 2 karakter, dan 3 karakter. Kata-kata tersebut tersebut biasanya adalah kata yang tidak memiliki makna, sebagai contoh 'e', 'yup', 'hah' dan lain sebagainya. *Preprocessing* data menghasilkan 46.605 data dengan 3 kolom, yaitu "no", "created_at", dan "text_preprocessed". Kolom "text_preprocessed" merupakan hasil proses *stopword removing*, *stemming*, dan penghapusan kata yang mengandung hanya 1, 2, dan karakter.

Potongan data hasil *preprocessing* ditunjukkan pada tabel 3.

Tabel 3. Hasil *Preprocessing* Data Opini

no	created_at	text_preprocessed
0	2020-08-09T23:43:01.000Z	polose dapat akun commerce bukalapak tokopedia shopee
1	2020-08-09T20:42:39.000Z	yups commerce belah
2	2020-08-09T18:23:50.000Z	beli commerce murah
3	2020-08-09T18:04:29.000Z	temu himalaya neem purifying mask tiptop depok harga murah commerce mell
4	2020-08-09T17:38:00.000Z	belanja diskon pacu kembang bisnis umkm commerce
...
46604	2020-10-27T01:17:50.000Z	buka toko oren buka gofud buka marketplace tiktok racun aman uang negara
46604	2020-10-27T00:51:56.000Z	cari marketplace facebook nder
46604	2020-10-27T00:46:48.000Z	benci facebook jual dimarketplace mana paham kali bantu retargeting card halaman facebook instagram iklan luar marketplace
46604	2020-10-27T00:15:43.000Z	kemarin bagi shopee store flash sale poco kali sharesave marketplace resmi xiaomi
46604	2020-10-26T23:56:37.000Z	belanja barang negeri imut lucu harga murah kualitas bagus gampang tinggal buka marketplace indonesia klik transaksi kirim negeri

Berdasarkan tabel 3, dilakukan pelabelan polaritas menggunakan kamus *lexicon* Senticnet 7 [19] dan persamaan (1). Hasil skor polaritas dari persamaan (1) digunakan untuk menentukan sentimen ulasan sesuai persamaan (2). Hasil persentase sentimen opini penggunaan *e-commerce* dan *marketplace* ditunjukkan pada gambar 2.



Gambar 2. Hasil Persentase Sentimen Opini

Sesuai gambar 2, mayoritas opini pada Twitter tentang penggunaan *e-commerce* dan *marketplace* bersifat negatif dengan jumlah persentase 80,3% atau 37,284 ulasan. Sementara ulasan positif berjumlah 18% atau 8,389 opini, sedangkan

sentimen netral hanya 1,7% atau 792 ulasan. Kata-kata yang sering muncul dalam sentimen positif adalah 1) *commerce*, 2) pakai, 3) langsung, 4) harga, dan 5) produk. Sementara kata yang sering muncul dalam sentimen negatif adalah 1) *commerce*, 2) beli, 3) langsung, 4) bayar, dan 5) beli.

Setelah ulasan diklasifikasikan berdasarkan jenis sentimennya, dilakukan proses pemilihan *feature candidate*. *Feature candidate* pada tiap kelompok opini sentimen positif dan negatif dipilih sesuai label kata benda atau *noun singular* (NN), *noun plural* (NNS), *proper noun singular* (NNP) dan *proper noun plural* (NNPS). Tiap kata yang dipilih sebagai *feature candidate* kemudian diberikan bobot menggunakan *Term Frequency* (TF) dan *Inverse Document Frequency* (IDF) menggunakan persamaan (3), (4) dan (5). Hasil 44 kata pada sentimen positif dan negatif dengan ranking tertinggi ditunjukkan pada tabel 4, dan 5.

Tabel 4. Hasil Perangkingan Kata

No	Kata	Nilai TF-IDF
1	commerce	746.7082691
2	marketplace	777.8205004
3	jual	515.7571084
4	harga	462.0391504
5	produk	371.6517989
...
40	administrator	118.8593002
41	manfaat	116.7069767
42	minat	130.7770241
43	layanan	97.96926731
44	dana	128.7887577

Tiap kata dari hasil nilai TF-IDF pada tabel 4 dilakukan perhitungan perbandingan kata menggunakan *Word2Vec*. Penggunaan *word embedding* *Word2Vec* mampu menghasilkan ekstraksi aspek yang lebih baik [26]. Library yang digunakan dalam proses ini adalah Gensim. Dalam proses *Word2Vec* digunakan data korpus Wikipedia Bahasa Indonesia. Karena data korpus berbentuk XML sehingga perlu diolah ke dalam bentuk teks. Proses pengolahan data korpus Wikipedia Bahasa Indonesia ke dalam bentuk teks ditunjukkan pada gambar 3.

```

2022-11-18 16:53:45,822: INFO: Saved 220000 articles
2022-11-18 16:53:48,183: INFO: Saved 230000 articles
2022-11-18 16:53:50,312: INFO: Saved 240000 articles
2022-11-18 16:53:52,973: INFO: Saved 250000 articles
2022-11-18 16:54:03,753: INFO: Saved 260000 articles
2022-11-18 16:54:12,488: INFO: Saved 270000 articles
2022-11-18 16:54:21,712: INFO: Saved 280000 articles
2022-11-18 16:54:31,762: INFO: Saved 290000 articles
2022-11-18 16:54:41,702: INFO: Saved 300000 articles
2022-11-18 16:54:51,838: INFO: Saved 310000 articles
2022-11-18 16:55:18,034: INFO: Saved 320000 articles
2022-11-18 16:55:29,433: INFO: Saved 330000 articles
2022-11-18 16:55:39,622: INFO: Saved 340000 articles
2022-11-18 16:55:51,125: INFO: Saved 350000 articles
2022-11-18 16:56:01,534: INFO: Saved 360000 articles
2022-11-18 16:56:12,132: INFO: Saved 370000 articles
2022-11-18 16:56:27,973: INFO: Saved 380000 articles
2022-11-18 16:56:37,787: INFO: Saved 390000 articles
2022-11-18 16:56:47,892: INFO: Saved 400000 articles
2022-11-18 16:56:56,992: INFO: Saved 410000 articles
2022-11-18 16:57:07,292: INFO: Saved 420000 articles
2022-11-18 16:57:16,962: INFO: Saved 430000 articles
2022-11-18 16:57:25,882: INFO: Saved 440000 articles
2022-11-18 16:57:34,682: INFO: Saved 450000 articles
2022-11-18 16:57:50,102: INFO: Saved 460000 articles
2022-11-18 16:58:02,702: INFO: finished iterating over Wikipedia
2004 articles, 143711344 positions before pruning articles short
2022-11-18 16:58:02,729: INFO: Finished Saved 469623 articles

```

Gambar 3. Hasil Konversi Korpus

Terdapat 469.623 artikel di dalam korpus Wikipedia Bahasa Indonesia sesuai gambar 3. Setelah data berhasil terkonversi, selanjutnya dilakukan training Word2Vec. Penelitian ini menggunakan beberapa parameter berbeda pada model Word2Vec, yaitu *windows size* dan *vector dimension*. Parameter *windows size* pada penelitian ini adalah 3, 6, dan 9. Sementara *vector dimension* yaitu 50, 150, dan 300. Sedangkan arsitektur Word2Vec adalah *Continuous Bag of Words* (CBOW). Pada domain tertentu, arsitektur CBOW memiliki hasil akurasi yang lebih baik daripada *Skip-gram*, lebih lanjut penelitian yang dilakukan oleh [27] menunjukkan dengan ukuran *word vector* 300 dicapai hasil akurasi tertinggi. Hasil kemiripan kata pada sentimen positif ditunjukkan pada tabel 5.

Tabel 5. Hasil Kemiripan Antar Kata Sentimen Positif

Pasangan Kata	Word2Vec		
	Windows Size		
	3	6	9
<i>commerce-marketplace</i>	0.847	0.856	0.781
<i>marketplace-transaksi</i>	0.512	0.531	0.438
harga-promosi	0.244	0.223	0.168
produk-toko	0.405	0.391	0.325
<i>barang-marketplace</i>	0.342	0.413	0.335
<i>toko-commerce</i>	0.389	0.349	0.331
promosi-media	0.303	0.298	0.276
bisnis-layanan	0.576	0.575	0.472
paket-kirim	0.283	0.283	0.283
<i>online-temu</i>	0.058	0.011	0.009

Ukuran *windows size* dan *vector dimension* yang lebih besar tidak selalu dapat meningkatkan nilai kemiripan antar kata sesuai tabel 5. Sebagai contoh pada pasangan kata promosi-media terjadi penurunan nilai kemiripan 1,85% pada *windows size* 6, dan kembali mengalami pengurangan nilai kemiripan pada *windows size* 9 sebesar 7,43%.

Sementara kenaikan nilai kemiripan ada pada pasangan kata *barang-commerce-marketplace*, *marketplace-transaksi*, dan *marketplace* dengan kenaikan sebesar. Sementara hasil kemiripan antar kata pada sentimen negatif ditunjukkan pada tabel 6.

Tabel 6. Hasil Kemiripan Antar Kata Sentimen Negatif

Pasangan Kata	Word2Vec		
	Windows Size		
	3	6	9
<i>commerce-marketplace</i>	0.781	0.848	0.759
<i>marketplace-transaksi</i>	0.438	0.521	0.454
harga-promosi	0.168	0.226	0.187
produk-toko	0.325	0.409	0.337
<i>barang-marketplace</i>	0.335	0.370	0.307
<i>toko-commerce</i>	0.331	0.359	0.302
promosi-media	0.276	0.313	0.288
bisnis-layanan	0.472	0.562	0.472
paket-kirim	0.283	0.048	0.289
<i>online-temu</i>	0.009	0.022	0.059

Secara keseluruhan kemiripan antar kata pada sentimen negatif mengalami peningkatan ketika menggunakan Word2Vec dengan *windows size* 6. Pada tabel 6, kenaikan nilai kemiripan antar kata sekitar 21,79%. Penambahan tertinggi pada pasangan kata harga-promosi dengan penambahan nilai kemiripan 34,39% pada *Windows Size* 6. Sementara pengurangan nilai kemiripan terjadi pada pasangan kata kirim-paket dengan nilai 83,02%. Sama halnya dengan hasil kemiripan kata pada sentimen positif, penggunaan *Windows Size* 9 dan *Vector Dimension* 300 menyebabkan pengurangan nilai kemiripan antar kata pada semua pasangan kata. Hal ini membuktikan bahwa ukuran dimensi vektor tidak berpengaruh peningkatan nilai kemiripan antar kata [27].

Nilai kedekatan makna yang dihasilkan oleh perhitungan *Cosine Similarity* pada model Word2Vec dievaluasi berdasarkan nilai korelasinya dengan nilai kemiripan pada *gold standard* Wordnet. Tujuan perhitungan nilai korelasi untuk melihat keberhasilan algoritma Word2Vec dalam melakukan perhitungan kesamaan antar kata. Hasil korelasi berdasarkan perhitungan *Pearson Correlation* pada persamaan (7) dapat dilihat pada tabel 7.

Tabel 7. Korelasi Word2Vec dan Wordnet

Windows Size	WordNet
3	0.438
6	0.446
9	0.596

Tabel 7 memperlihatkan nilai korelasi (r) nilai kesamaan antar kata yang diproses menggunakan Word2Vec dan *gold standard* Wordnet. Penggunaan

Windows Size yang lebih besar mempengaruhi nilai *pearson correlation*. Pada implementasi *Windows Size* 6 terjadi peningkatan nilai korelasi (*r*) sebesar 1,77%, sementara pada *Windows Size* 9 nilai *pearson correlation* meningkat sebesar 35,91%. Contoh perbandingan kemiripan antar kata antara Word2Vec dan *gold standard* ditunjukkan pada tabel 8. Sesuai tabel 8, semua kata yang menjadi kandidat aspek terlebih dahulu dicari kosakatanya dalam bahasa Inggris. Sebagai contoh kata "media", kosakata dalam bahasa Inggris menjadi "medium.n.01". Contoh lainnya adalah kata "produk", kata yang sesuai dalam bahasa Inggris adalah "commodity.n.01". Peralihan kata dari bahasa Indonesia ke dalam bahasa Inggris pada Wordnet disertai dengan kelas katanya. Pada kata "commodity.n.01", huruf "n" menunjukkan kelas kata tersebut adalah nomina (kata benda). Kesamaan antar kata dihitung berdasarkan

path_based_similarity, yaitu merupakan jarak *path* antar kedua kata.

Berdasarkan kemiripan antara kata, selanjutnya dapat ditentukan aspek-aspek di dalam ABSA penggunaan *e-commerce* dan *marketplace*. Aspek-aspek tersebut didapatkan melalui perangkingan berdasarkan nilai kemiripan Word2Vec pada *windows size* 6 dan *vector dimension* 150. Nilai tersebut dipilih karena adanya penambahan nilai *similarity* antar kata. Selanjutnya kata yang menjadi aspek ditentukan berdasarkan nilai kemiripan dua kata yang diklasifikasikan pada korelasi sedang dan korelasi kuat (tabel 1). Hasil klasifikasi terdapat 14 aspek seperti ditampilkan pada tabel 9. Sama halnya dengan penentuan aspek, kata-kata di dalam aspek ditentukan sesuai nilai kemiripan antar kata yang diklasifikasikan pada korelasi sedang dan korelasi kuat.

Tabel 8. Beberapa Perbandingan Kemiripan Antar Kata

No.	Kata	Synset Words	Windows Size					
			3		6		9	
			Wordnet	Word2Vec	Wordnet	Word2Vec	Wordnet	Word2Vec
1	commerce-marketplace	commerce.n.01-market.n.01	0.167	0.847	0.167	0.856	0.167	0.781
2	commerce-jual	commerce.n.01-sell.n.01	0.250	0.371	0.250	0.407	0.250	0.356
3	marketplace-jual	market.n.01-sell.n.01	0.111	0.355	0.111	0.425	0.111	0.400
4	marketplace-online	market.n.01-on-line.a.01	0.000	0.587	0.000	0.557	0.000	0.499
5	marketplace-toko	market.n.01-retailing.n.01	0.125	0.391	0.125	0.430	0.125	0.329
6	jual-harga	sell.n.01-price.n.04	0.077	0.525	0.077	0.540	0.077	0.441
7	jual-produk	sell.n.01-commodity.n.01	0.067	0.442	0.067	0.444	0.067	0.372
8	jual-laku	sell.n.01-salable.a.01	0.000	0.260	0.000	0.302	0.000	0.239
9	produk-jual	commodity.n.01-sell.n.01	0.067	0.442	0.067	0.444	0.067	0.372
10	produk-harga	commodity.n.01-price.n.04	0.091	0.548	0.091	0.552	0.091	0.457
11	produk-manfaat	commodity.n.01-profit.v.01	0.000	0.480	0.000	0.454	0.000	0.373
12	bisnis-commerce	occupation.n.01-commerce.n.01	0.167	0.618	0.167	0.620	0.167	0.586
13	bisnis-marketplace	occupation.n.01-market.n.01	0.333	0.540	0.333	0.549	0.333	0.483
14	bisnis-manfaat	occupation.n.01-profit.v.01	0.000	0.354	0.000	0.339	0.000	0.255
15	media-commerce	medium.n.01-commerce.n.01	0.077	0.511	0.077	0.496	0.077	0.398
16	media-marketplace	medium.n.01-market.n.01	0.077	0.472	0.077	0.458	0.077	0.350
17	media-platform	medium.n.01-platform.n.01	0.167	0.518	0.167	0.555	0.167	0.441
18	layanan-promosi	service.n.15-discount.n.01	0.100	0.410	0.100	0.407	0.100	0.318
19	layanan-manfaat	service.n.15-profit.v.01	0.000	0.420	0.000	0.425	0.000	0.343
20	layanan-platform	service.n.15-platform.n.01	0.063	0.705	0.063	0.695	0.063	0.578

Tabel 9. Aspek dan Fitur Aspek

No	Aspek	Fitur Aspek
1	commerce	marketplace, produk, online, transaksi, bisnis, platform, digital, layanan
2	marketplace	commerce, produk, online, transaksi, bisnis, pasar, paket, platform, order, digital, layanan
3	jual	harga, transaksi, barang, belanja, beli
4	harga	jual, produk, transaksi, barang, rupiah, ongkos, pasar, uang
5	bisnis	commerce, marketplace, produk, transaksi, toko, belanja, pasar, layanan
6	pasar	marketplace, harga, produk, transaksi, toko, bisnis, belanja, layanan
7	paket	marketplace, produk, transaksi, platform, layanan
8	juta	rupiah, ribu
9	platform	commerce, marketplace, produk, media, paket, digital, layanan
10	beli	jual
11	ribu	rupiah, juta
12	uang	harga, transaksi, barang, rupiah, dana
13	layanan	commerce, marketplace, produk, transaksi, bisnis, pasar, paket, jasa, platform, digital
14	dana	ongkos, uang, manfaat

Terdapat kata-kata yang diklasifikasikan sebagai fitur aspek pada beberapa aspek sebagaimana ditampilkan pada tabel 9. Misalkan kata “commerce” dikategorikan pada aspek marketplace, bisnis, platform serta layanan. Contoh lainnya adalah kata “platform” yang menjadi fitur pada aspek commerce, marketplace, paket, dan layanan. Aspek fitur yang menjadi bagian beberapa fitur akan ditentukan fiturnya berdasarkan kedekatan nilai kemiripannya. Sebagai contoh, berdasarkan nilai kemiripan kata “commerce” dengan kata lainnya. Selain menjadi fitur, kata “commerce” ditentukan menjadi fitur “marketplace”, karena memiliki nilai kata “commerce” dan “marketplace” memiliki nilai kemiripan tertinggi dibandingkan dengan kata lainnya, yaitu 0,855.

4. DISKUSI

Ekstraksi aspek opini penggunaan aplikasi e-commerce dan marketplace dilakukan dengan memilih kata berdasarkan Part of Speech (POS) Tagging. Penggunaan POS Tagging di dalam Word2Vec dapat mengenali kata yang memiliki kemiripan makna sesuai kelas katanya [28]. Selain itu, penggunaan POS Tagging dapat meningkatkan akurasi dalam Aspect-Based Sentiment Analysis serta mampu mengekstraksi multi aspek melalui kombinasi Type Dependency Relations (TDR) yang tepat [29]. Implementasi windows size dan vector dimension terbukti dapat meningkatkan nilai kemiripan kata yang memiliki kedekatan makna mencapai 34,39% (tabel 5).

Kedekatan makna kata hasil Word2Vec dibandingkan dengan kemiripan kata menggunakan path_similarity berdasarkan kamus WordNet. Hasil kedekatan makna antara kata yang diperoleh dari

cosine_similarity Word2Vec dan path_similarity WordNet dicari nilai korelasinya. Nilai korelasi tertinggi didapatkan pada percobaan Word2Vec menggunakan windows size 9 dan vector dimension 300, yaitu adanya korelasi sedang. Sedangkan uji coba dengan windows size 3, 6, dan vector dimension 50, serta 150 hanya berkorelasi rendah dengan hasil path_similarity WordNet. Salah satu penyebab hubungan antara hasil Word2Vec dan WordNet pada posisi sedang atau bernilai 0,596 adalah dipengaruhi oleh perbedaan kelas kata yang didapatkan dari proses POS Tagging ketika diubah ke dalam kosakata WordNet. Kata “marketplace” ditandai sebagai kata benda (noun), sedangkan kata yang dibandingkan, yaitu “online” merupakan kata sifat (adjective). Karena perbedaan kelas antar dua kata tersebut, maka kedekatan antar kata tidak bisa diukur (tabel 8).

Sebagaimana ditunjukkan pada tabel 9. Aspek yang berhasil diekstraksi berjumlah 14. Fitur didapatkan berdasarkan nilai kemiripan suatu kata dengan kata lainnya dengan rentang $0,5 \geq$ nilai kemiripan < 1 pada Word2Vec dengan windows size 6 dan vector dimension 150.

5. KESIMPULAN

Ekstraksi aspek dapat dilakukan melalui proses POS Tagging untuk memilih kata yang memiliki label label kata benda atau noun singular (NN), noun plural (NNS), proper noun singular (NNP) dan proper noun plural (NNPS) sebagai kandidat aspek, sementara penentuan aspek dilakukan berdasarkan pembobotan TF-IDF. Untuk meningkatkan nilai kemiripan antar kata digunakan word embedding Word2Vec dengan arsitektur CBOW. Sebanyak 44 kata dengan bobot tertinggi diproses pada Word2Vec. Ukuran windows size diatur 6 dan vector dimension adalah 150, nilai tersebut terbukti menghasilkan nilai kemiripan antar kata tertinggi. Sesuai nilai kemiripan antar kata ditentukan aspek dan kata-kata yang menjadi fitur suatu aspek. Sebanyak 14 aspek dipilih berdasarkan nilai kedekatan makna dua kata dengan korelasi sedang dan korelasi kuat, selain itu juga digunakan untuk mengelompokkan fitur pada suatu aspek.

Meskipun berhasil mengekstraksi aspek berdasarkan opini penggunaan aplikasi e-commerce dan marketplace. Namun penelitian ini masih memiliki kekurangan, yaitu adanya perbedaan pelabelan kelas kata antara hasil POS Tagging dan WordNet sehingga tidak dapat dilakukan pengukuran kedekatan antar kata. Keterbatasan lainnya adalah pada proses normalisasi kata-kata opini yang tidak terstruktur dan cenderung menggunakan bahasa tidak baku.

Sebagai saran untuk penelitian kedepan. Ekstraksi aspek dapat ditingkat melalui pendekatan machine learning untuk proses normalisasi kata sehingga kata-kata yang tidak baku dapat diproses menjadi kata baku. Selain itu ekstraksi fitur tidak

hanya melalui pelabelan POS *Tagging* saja, tetapi bisa ditingkatkan dengan menggunakan *Dependency Tree Representation* untuk melihat hubungan antar kata di dalam suatu kalimat.

UCAPAN TERIMA KASIH

Ucapan terima kasih disampaikan kepada Direktorat Riset, Teknologi, dan Pengabdian Masyarakat Direktorat Jenderal Pendidikan Tinggi, Riset, dan Teknologi Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi Republik Indonesia yang telah mendanai penelitian ini melalui skema Penelitian Dosen Pemula.

DAFTAR PUSTAKA

- [1] We Are Social, "Digital 2022 Indonesia, February 2022," We Are Social, New York, 2022.
- [2] Badan Pusat Statistik, "Analisis Hasil Survei Dampak COVID-19 terhadap Pelaku usaha," 2020.
- [3] F. Xu, Z. Pan, and R. Xia, "E-commerce product review sentiment classification based on a naïve Bayes continuous learning framework," *Inf. Process. Manag.*, vol. 57, no. 5, p. 102221, 2020, doi: 10.1016/j.ipm.2020.102221.
- [4] A. E. E. Sobaih, I. A. Palla, and A. Baquee, "Social Media Use in E-Learning amid COVID 19 Pandemic: Indian Students' Perspective," *Int. J. Environ. Res. Public Health*, vol. 19, no. 9, pp. 1–12, 2022, doi: 10.3390/ijerph19095380.
- [5] J. Mou and M. Benyoucef, "Consumer behavior in social commerce: Results from a meta-analysis," *Technol. Forecast. Soc. Change*, vol. 167, no. March, p. 120734, 2021, doi: 10.1016/j.techfore.2021.120734.
- [6] APJII, "Profil Internet Indonesia 2022," 2022.
- [7] C. Zachlod, O. Samuel, A. Ochsner, and S. Werthmüller, "Analytics of social media data – State of characteristics and application," *J. Bus. Res.*, vol. 144, no. January, pp. 1064–1076, 2022, doi: 10.1016/j.jbusres.2022.02.016.
- [8] B. Ozyurt and M. A. Akcayol, "A new topic modeling based approach for aspect extraction in aspect based sentiment analysis: SS-LDA," *Expert Syst. Appl.*, vol. 168, no. November 2020, p. 114231, 2021, doi: 10.1016/j.eswa.2020.114231.
- [9] Ł. Augustyniak, T. Kajdanowicz, and P. Kazienko, "Comprehensive analysis of aspect term extraction methods using various text embeddings," *Comput. Speech Lang.*, vol. 69, 2021, doi: 10.1016/j.csl.2021.101217.
- [10] Q. Liu, J. Wang, D. Zhang, Y. Yang, and N. Wang, "Text features extraction based on tf-idf associating semantic," in *2018 IEEE 4th International Conference on Computer and Communications, ICC 2018*, 2018, pp. 2338–2343, doi: 10.1109/CompComm.2018.8780663.
- [11] N. Passalis and A. Tefas, "Learning bag-of-embedded-words representations for textual information retrieval," *Pattern Recognit.*, vol. 81, pp. 254–267, 2018, doi: 10.1016/j.patcog.2018.04.008.
- [12] D. Yan, K. Li, S. Gu, and L. Yang, "Network-Based Bag-of-Words Model for Text Classification," *IEEE Access*, vol. 8, pp. 82641–82652, 2020, doi: 10.1109/ACCESS.2020.2991074.
- [13] M. Hamka, N. Alfatari, and D. R. Sari, "Analisis Sentimen Produk Kecantikan Jenis Serum Menggunakan Algoritma Naïve Bayes Classifier," *J. Sist. Komput. dan Inform.*, vol. 4, no. September, pp. 64–71, 2022, doi: 10.30865/json.v4i1.4740.
- [14] B. Bansal and S. Srivastava, "Sentiment classification of online consumer reviews using word vector representations," *Procedia Comput. Sci.*, vol. 132, pp. 1147–1153, 2018, doi: 10.1016/j.procs.2018.05.029.
- [15] V. K. Ayyadevara, *Pro Machine Learning Algorithms*. Hyderabad: Apress, 2018.
- [16] I. Yunanto and S. Yulianto, "Twitter Sentiment Analysis Pedulilindungi Application Using Naïve Bayes and Support Vector Machine Analisis Sentimen Twitter Aplikasi Pedulilindungi," *J. Tek. Inform.*, vol. 3, no. 4, pp. 807–814, 2022, doi: 10.20884/1.jutif.2022.3.4.292.
- [17] R. Cahyadi, A. Damayanti, and D. Aryadani, "Recurrent Neural Network (RNN) dengan Long Short Term Memory (LSTM) untuk Analisis Sentimen data instagram," *J. Inform. dan Komput.*, vol. 5, no. 1, pp. 1–9, 2020, doi: 10.26798/jiko.v5i1.407.
- [18] Y. N. Prasetya, D. Winarso, and Syahril, "Penerapan Lexicon Based Untuk Analisis Sentimen Pada Twiter Terhadap Isu Covid-19," *FASILKOM*, vol. 11, no. 2, pp. 97–103, 2021.

- [19] E. Cambria, Q. Liu, S. Decherchi, F. Xing, and K. Kwok, "SenticNet 7: A Commonsense-based Neurosymbolic AI Framework for Explainable Sentiment Analysis," in *Proceedings of the Language Resources and Evaluation Conference*, 2022, no. June, pp. 3829–3839.
- [20] E. W. Pamungkas and D. G. P. Putri, "An experimental study of lexicon-based sentiment analysis on Bahasa Indonesia," in *Proceedings - 2016 6th International Annual Engineering Seminar, InAES 2016*, 2017, pp. 28–31, doi: 10.1109/INAES.2016.7821901.
- [21] L. Vu and T. Le, "A lexicon-based method for Sentiment Analysis using social network data," in *Int'l Conf. Information and Knowledge Engineering*, 2017, no. September, pp. 10–16.
- [22] W. Yulita, E. D. Nugroho, and M. H. Algifari, "Analisis Sentimen Terhadap Opini Masyarakat Tentang Vaksin Covid - 19 Menggunakan Algoritma Naïve Bayes Classifier," *JDMSI*, vol. 2, no. 2, pp. 1–9, 2021.
- [23] L. N. T. Manalu, M. A. Bijaksana, and A. A. Suryani, "Analysis of the Word2Vec model for semantic similarities in Indonesian words," *2019 7th Int. Conf. Inf. Commun. Technol. ICoICT 2019*, 2019, doi: 10.1109/ICoICT.2019.8835330.
- [24] D. Jatnika, M. A. Bijaksana, and A. A. Suryani, "Word2vec model analysis for semantic similarities in English words," *Procedia Comput. Sci.*, vol. 157, pp. 160–167, 2019, doi: 10.1016/j.procs.2019.08.153.
- [25] A. Elekes, M. Schaefer, and K. Boehm, "On the Various Semantics of Similarity in Word Embedding Models," *Proc. ACM/IEEE Jt. Conf. Digit. Libr.*, 2017, doi: 10.1109/JCDL.2017.7991568.
- [26] J. Z. Maitama, N. Idris, A. Abdi, L. Shuib, and R. Fauzi, "A systematic review on implicit and explicit aspect extraction in sentiment analysis," *IEEE Access*, vol. 8, pp. 194166–194191, 2020, doi: 10.1109/ACCESS.2020.3031217.
- [27] P. Michelle, F. Z. Ruskanda, and A. Purwarianti, "Development of Domain-Specific Lexicon for Aspect-Based Sentiment Analysis," *2020 7th Int. Conf. Adv. Informatics Concepts, Theory Appl. ICAICTA 2020*, 2020, doi: 10.1109/ICAICTA49861.2020.9429059.
- [28] D. Suleiman and A. A. Awajan, "Using part of speech tagging for improving word2vec model," *2019 2nd Int. Conf. New Trends Comput. Sci. ICTCS 2019 - Proc.*, 2019, doi: 10.1109/ICTCS.2019.8923081.
- [29] A. S. Shafie, N. M. Sharef, M. A. A. Murad, and A. Azman, "Aspect Extraction Performance with POS Tag Pattern of Dependency Relation in Aspect-based Sentiment Analysis," *Proc. - 2018 4th Int. Conf. Inf. Retr. Knowl. Manag. Diving into Data Sci. CAMP 2018*, pp. 107–112, 2018, doi: 10.1109/INFRKM.2018.8464692.