

## ANALYSIS OF MARKETPLACE CONVERSATION TRENDS ON TWITTER PLATFORM USING K-MEANS

Ulil Amri Nasron<sup>1</sup>, Muhammad Habibi<sup>2</sup>

Program Studi Informatika, Fakultas Teknik dan Teknologi Informasi  
Universitas Jenderal Achmad Yani Yogyakarta  
Jl. Siliwangi, Ringroad Barat, Banyuraden, Gamping, Sleman  
Daerah Istimewa Yogyakarta 55293  
Email: <sup>1</sup>ulilamri69@gmail.com, <sup>2</sup>muhammadhabibi17@gmail.com

### Abstract

*Businesses began to switch from the process of marketing products using conventional media to the process of marketing products using internet and social media. The switch of businesspeople to the process of marketing products through the internet and social media is because the marketing costs are cheaper than using conventional media. The problem that is often faced by business people when marketing products on social media is that they rarely see a marketplace that is becoming a trend and is being discussed by consumers on social media, so the marketing process is carried out less than the maximum. This study aims to analyze conversation trends related to the marketplace on the social media platform Twitter. The method used in this study is the K-Means Clustering method. This research succeeded in analyzing conversation trends related to the marketplace so that it can be used as a basis for consideration of businesspeople in choosing a marketplace. Marketplace trend analysis results show that Shopee, Lazada, and Tokopedia are marketplaces that are the subject of conversation on the social media platform Twitter.*

**Keyword:** Marketplace, Text Mining, K-Means, Twitter, Clustering.

### 1. Latar Belakang

Analisis tren adalah suatu analisis yang menggambarkan atau menunjukkan perubahan rata-rata suatu variabel tertentu dari waktu ke waktu. Perubahan rata-rata suatu variabel yang mengalami kecenderungan penurunan nilai disebut tren negatif, sedangkan perubahan rata-rata suatu variabel yang mengalami kecenderungan peningkatan disebut tren positif [1]. Berdasarkan observasi, terdapat beberapa pelaku bisnis yang mengalami kesulitan dalam menentukan pemilihan *marketplace* saat mereka mulai beralih dari proses penjualan dengan cara konvensional ke proses penjualan melalui media sosial dan internet. Beralihnya pelaku bisnis ke proses penjualan melalui internet dan media sosial disebabkan karena biaya pemasaran yang digunakan lebih murah dibandingkan dengan menggunakan media konvensional. Permasalahan yang sering dihadapi pelaku bisnis saat melakukan pemasaran di media sosial adalah mereka jarang melihat apa yang sedang konsumen perbincangkan mengenai produk yang ingin mereka pasarkan, serta mereka jarang melihat *marketplace* yang sedang menjadi tren dan diperbincangkan oleh konsumen, sehingga produk yang mereka jual hanya terjual sedikit atau bahkan tidak ada yang terjual samasekali. Pembahasan terkait tren *marketplace* banyak diperbincangkan di media sosial, salah satunya di *platform* Twitter.

Twitter merupakan salah satu media sosial yang banyak digunakan oleh masyarakat dan populer di Indonesia. Hal ini berdasarkan berdasarkan pada hasil riset Wearesosial Hootsuite yang dirilis pada bulan Januari 2019 pengguna media sosial di Indonesia mencapai 150 juta atau sebesar 56% dari total populasi. Jumlah tersebut naik 20% dari survei sebelumnya. Sementara pengguna media sosial *mobile* (ponsel) mencapai 130 juta atau 48% dari total populasi [2]. Sedangkan berdasarkan data dari kompas.com, pengguna harian Twitter pada kuartal ke-3 tahun 2019 meningkat 17% ke angka 145 juta pengguna [3]. Pengguna Twitter dapat membuat pesan-pesan pendek yang disebut dengan *tweet*, dimana melalui *tweet* tersebut, pengguna Twitter dapat saling berhubungan, berbagi pendapat, dan menemukan berbagai informasi dari seluruh dunia.

Analisis tren di dalam media sosial Twitter menjadi penting bagi pelaku bisnis untuk mengetahui pola kecenderungan konsumen terhadap produk barang atau jasa yang dimiliki. Dalam hal ini, *tweet*/kicauan yang ada di dalam Twitter menjadi penting untuk dianalisis dan diterjemahkan menjadi sesuatu yang lebih bermakna sehingga dapat dimanfaatkan oleh pelaku bisnis. Dalam dunia bisnis, pelaku bisnis membutuhkan analisis tren untuk meningkatkan efektivitas dalam melakukan pemasaran, dan *forecasting*.

Berdasarkan permasalahan tersebut maka diperlukan model untuk melakukan analisis percakapan tren *marketplace* pada Twitter. Penelitian yang menggunakan data media sosial Twitter sudah banyak dilakukan, diantaranya adalah penelitian analisis konten jejaring sosial Twitter dalam kasus pemilihan Gubernur DKI Jakarta 2017 [4] serta model *monitoring* sebaran penyakit demam berdarah di Indonesia berdasarkan analisis pesan Twitter [5]. Penelitian ini menggunakan metode K-Means sebagai solusi untuk pengklasifikasian dari sebuah objek. Metode K-Means banyak digunakan dalam penelitian terkait dengan pengklasteran data, diantaranya adalah penelitian terkait klasterisasi *user* karakteristik berdasarkan *hashtag* pada *platform* Instagram [6], klasterisasi tipe pembelajar sebagai parameter evaluasi kualitas pendidikan di perguruan tinggi [7], pengelompokan kunjungan wisatawan ke objek wisata unggulan di Prov. DKI Jakarta [8], serta digunakan untuk kuantisasi foto peradangan pada kulit [9]. Penerapan *term frequency* dapat dilakukan untuk mengekstraksi teks dari komentar mahasiswa terhadap dosen [10]. TF-IDF (*Term frequency- Inverse Document Frequency*) merupakan *metric* yang umum digunakan dalam proses kategorisasi teks [11]. TF-IDF terdiri dari dua buah nilai komponen yaitu *term-frequency* dan *inverse document frequency*. Penggunaan TF-IDF dapat dengan baik pada metode penambangan teks [12].

## 2. Metodologi Penelitian

Penelitian ini dilaksanakan dalam empat tahapan. Tahapan tersebut diantaranya *collecting data*, *preprocessing*, *feature selection* dan *Clustering*. Tahapan dalam penelitian ini dapat dilihat pada Gambar 1.

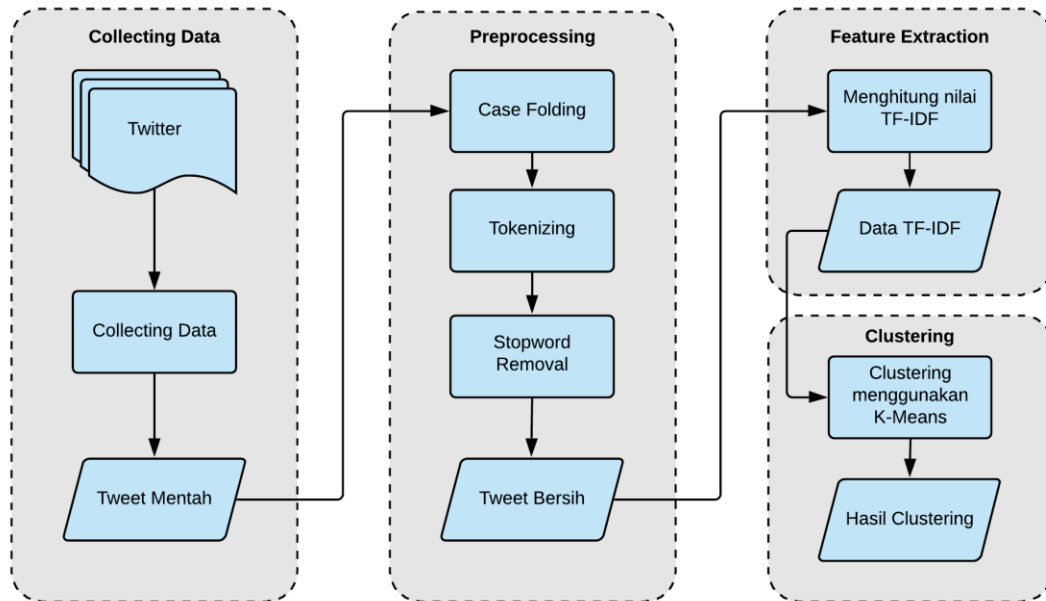
### 2.1. Collecting Data

Penelitian ini menggunakan data *tweet* dan *re-tweet* yang terkait dengan *marketplace* yang terdapat di Indonesia. Data yang diambil adalah data *tweet* dan *re-tweet* terkait dengan hastag dari Bukalapak, Lazada, Shopee, dan Tokopedia yang berupa #bukalapak, #lazada, #shopee, dan #tokopedia. Lama waktu yang digunakan dalam mengumpulkan data adalah selama 20 hari, dimulai dari tanggal 23 Februari – 14 Maret 2019, dengan jumlah data *tweet* dan *re-tweet* yang diperoleh sebanyak 7124.

## 2.2. Preprocessing

*Text preprocessing* merupakan tahapan awal dari *Text Mining*, tahap ini meliputi proses untuk mempersiapkan data tekstual yang akan digunakan agar dapat diproses pada tahapan berikutnya [13]. Proses yang dilakukan pada tahapan ini antara lain:

1. *Case folding* adalah proses konversi teks menjadi satu huruf besar atau kecil.
2. *Tokenizing* adalah proses untuk membagi teks ke dalam token.
3. *Stop-word removal* adalah proses menghilangkan kata-kata dari dokumen yang tidak mempunyai peran penting dalam memberi pola atau informasi.



Gambar 1. Tahapan Penelitian

## 2.3. Feature Extraction

*Feature extraction* atau ekstraksi fitur merupakan proses ekstraksi untuk mengidentifikasi entitas-entitas yang dimaksud [14]. Tahapan ini melakukan pencarian nilai *Term frequency* (TF) dan IDF (*Inverse Document Frequency*) pada data *text* yang sudah dikumpulkan. Dengan tujuan melihat seberapa penting kemunculan token pada suatu korpus. Adapun perhitungan nilai bobot TF-IDF diperoleh dari persamaan (1) berikut [15]:

$$w_{t,d} = tf_{t,d} \times \log \frac{N}{df_t} \quad (1)$$

keterangan:

$tf_{t,d}$  = merupakan bobot suatu *term t* pada dokumen *d*

$N$  = merupakan jumlah dokumen keseluruhan

$df_t$  = merupakan banyaknya dokumen yang memuat *term t*.

## 2.4. Clustering

*Clustering* merupakan penggunaan teknik data mining dimana sekelompok objek yang sama digabungkan bersama untuk membentuk *Cluster*, *Cluster* ini berbeda dari objek di *Cluster* lain [16]. Pada tahap ini digunakan K-Means untuk mengklusterisasi data twitter terkait *marketplace*. Adapun perhitungan K-Means diperoleh dari persamaan (2) berikut:

$$d(x_j, c_j) = \sqrt{\sum_{j=1}^n (x_j - c_j)^2} \quad (2)$$

Keterangan:

d = jarak

n = banyaknya objek

j = (dimulai dari 1 sampai n)

x<sub>j</sub> = *feature* objek ke j terhadap x

c<sub>j</sub> = *centroid feature* ke j

K-Means *clustering* merupakan suatu metode *data mining* yang melakukan proses pemodelan *unsupervised learning*, yaitu untuk melakukan prediksi atau klasifikasi tidak perlu dilatih atau proses *training* terlebih dahulu. Tujuan dari Algoritma K-Means *Clustering* adalah untuk mengelompokkan objek yang hampir sama dalam area tertentu.

Secara umum langkah Algoritma K-Means [7] adalah:

1. Menentukan banyaknya kluster (k)
2. Menentukan *centroid*
3. Analisis apakah *centroid* berubah berdasarkan means setiap *feature* objek (data)?
  - a. [ya], ubah *centroid* menjadi centroid baru
  - b. [tidak], selesai
4. Hitung *distance space* (jarak kedekatan) objek dengan *centroid*
5. Kelompokkan objek berdasarkan kedekatan objek dengan *centroid*

## 3. Analisis dan Pembahasan

### 3.1. Term frequency – Inverse Document Frequency (TF-IDF)

Pada bagian ini akan membahas hasil dari penelitian yang sudah dilakukan. Hasil dari TF dapat dilihat pada Tabel 1.

Tabel 1. *Term frequency*.

	sapa	promo	jual	lewat	diskon	tokopedia	...
1	1	0	0	0	0	1	...
2	0	0	1	1	0	3	...
3	0	0	0	0	0	2	...
4	0	1	0	0	0	1	...
5	0	3	0	0	0	1	...
6	0	1	0	0	0	1	...
7	0	0	1	1	0	3	...

8	0	1	0	0	2	2	...
9	0	0	0	0	0	1	...
10	0	0	1	1	0	3	...
...	...	...	...	...	...	...	...
TF	1	6	3	3	2	18	

Tabel 1. merupakan hasil *term frequency* dari corpus atau kamus kata yang sudah dikumpulkan. Misalkan dari corpus tersebut didapatkan hasil *frequency* dari *term* sapa, promo, jual, lewat, diskon dan tokopedia berturut-turut adalah 1, 6, 3, 3, 2, 18. Kemudian langkah berikutnya adalah menghitung nilai TF-IDF menggunakan persamaan (1). Hasil perhitungan nilai TF-IDF dari Tabel 1. dapat dilihat pada Tabel 2.

Tabel 2. TF-IDF.

	<b>sapa</b>	<b>promo</b>	<b>Jual</b>	<b>lewat</b>	<b>diskon</b>	<b>Tokopedia</b>	<b>...</b>
1	1	0	0	0	0	1	...
2	0	0	1	1	0	3	...
3	0	0	0	0	0	2	...
4	0	1	0	0	0	1	...
5	0	3	0	0	0	1	...
6	0	1	0	0	0	1	...
7	0	0	1	1	0	3	...
8	0	1	0	0	2	2	...
9	0	0	0	0	0	1	...
10	0	0	1	1	0	3	...
...	...	...	...	...	...	...	...
IDF	Log (10/1) = 1	Log (10/6) = 0.211	Log (10/3) = 0.522	Log (10/3) = 0.522	Log (10/2) = 0.698	Log (10/18) = - 0.255	...
<b>TF x IDF</b>							
1	1	0	0	0	0	- 0.255	...
2	0	0	0.522	0.522	0	- 0.765	...
3	0	0	0	0	0	- 0.51	...
4	0	0.211	0	0	0	- 0.255	...
5	0	0.633	0	0	0	- 0.255	...
6	0	0.211	0	0	0	- 0.255	...
7	0	0	0.522	0.522	0	- 0.765	...
8	0	0.211	0	0	0.698	- 0.51	...
9	0	0	0	0	0	- 0.255	...
10	0	0	0.522	0.522	0	- 0.51	...
...	...	...	...	...	...	...	...

### 3.2. Penentuan Jumlah Kluster Terbaik

Metode *Elbow* merupakan suatu metode yang digunakan untuk menghasilkan informasi dalam menentukan jumlah *cluster* terbaik dengan cara melihat persentase hasil perbandingan antara jumlah *cluster* yang akan membentuk siku pada suatu titik. Metode ini memberikan ide/gagasan dengan cara memilih nilai *cluster* dan kemudian menambah nilai *cluster* tersebut untuk dijadikan model data dalam penentuan *cluster* terbaik. [17].

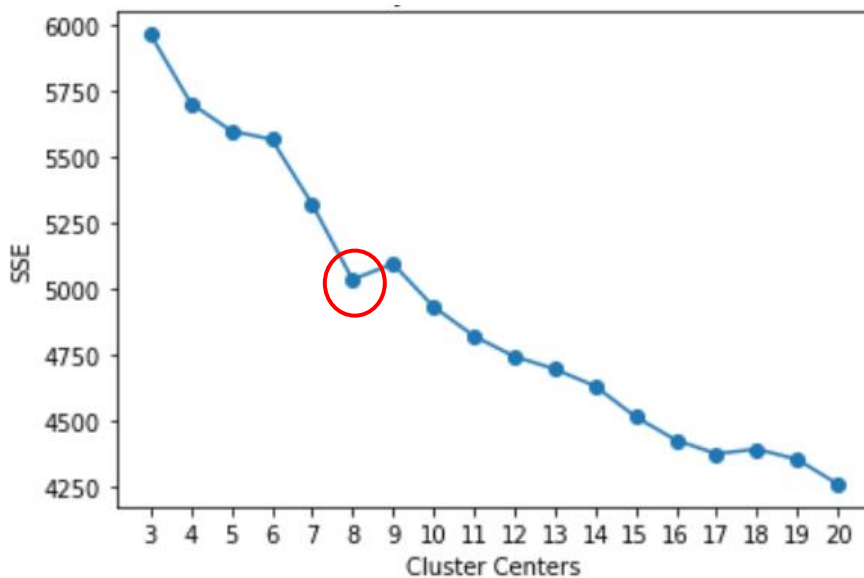
Untuk mendapatkan perbandingannya adalah dengan menghitung *Sum of Square Error* (SSE) dari masing-masing nilai *cluster*. Karena semakin besar jumlah *cluster*  $K$  maka nilai SSE akan semakin kecil. Untuk menghitung nilai SSE pada K-Means menggunakan persamaan (3) berikut.

$$SSE = \sum_{K=1}^K \sum_{x_i \in S_K} \|x_i - c_k\|_2^2 \quad (3)$$

Langkah penggunaan algoritma metode *Elbow* dalam menentukan nilai  $K$  pada K-Means adalah sebagai berikut:

1. Dimulai pada penentuan jumlah *cluster* 1.
2. Menginisialisasi nilai awal
3. Menaikkan nilai  $K$
4. Menghitung hasil *sum of square error* dari tiap nilai  $K$
5. Melihat hasil *sum of square error* dari nilai  $K$  yang turun secara drastis
6. Menetapkan nilai  $K$  yang beberbentuk siku
7. Selesai.

Berikut merupakan grafik hasil perhitungan jumlah *cluster* terbaik menggunakan metode *elbow*, dapat dilihat pada Gambar 2.



Gambar 2. Grafik metode Elbow.

Berdasarkan hasil perhitungam, dapat dilihat ada beberapa nilai K yang mengalami penurunan paling besar dan selanjutnya hasil dari nilai K akan turun secara perlahan-lahan sampai hasil dari nilai K tersebut stabil. Terlihat penurunan drastis membentuk siku pada titik K=8 maka nilai *cluster* k yang ideal adalah K=8. Dengan demikian akan diperoleh nilai K yang sesuai untuk digunakan dalam proses *clustering* adalah 8 *cluster*.

### 3.3. Clustering

Dalam penelitian ini data dikelompokkan ke dalam 8 *cluster* sesuai dengan hasil jumlah *cluster* terbaik. Contoh hasil *Clustering* data *tweet* dapat dilihat pada Tabel 3.

Tabel 3. Hasil *Clustering*.

<i>Users</i>	<i>Tweets</i>	<i>Cluster</i>
hamimth	Pak Menyapa jambi Kamis 14Maret2019 Promise Shopee Tokopedia property PrabowoMenyapaJambi indonesia Indonesian prabowo ShopeeID PramborsThrowback bum	5
DigiCrossNet	Jual Speaker Fenda T200X mujidigicross Tokopedia SpeakerFendaT200X Tokopedia lewat	6
kireta44	Iya Tokopedia LOL	6
y4n1_82	Jual wok pan power cook modern Fashafira Tokopedia wokpanpowercookmodern Tokopedia via	6
DigiCrossNet	Jual Epson Printer A3 L1300 Hitam Print mujidigicross Tokopedia EpsonPrinterA3L1300HitamPrint Tokopedia lewat	6
DigiCrossNet	Jual Epson Printer A3 L1300 Hitam Print mujidigicross Tokopedia EpsonPrinterA3L1300HitamPrint Tokopedia lewat	6
Trivoksel	Jual PDU 6 outlet 16A Surge Protection Import Racknpower Tokopedia PDU6outlet16ASurgeProtectionImport Tokopedia lewat	6
Trivoksel	Jual PDU 12 outlets 16A with Surge Protection Racknpower Tokopedia PDU12outlets16AwithSurgeProtection Tokopedia lewat	6
keiskeipusat	Keiskei Flash Sale di Tokopedia Jam 1200 Jangan Sampai Ketinggalan Yaaa Diskon 10 Untuk Produk Paket Cek Langsung keiskei penumbuhrambut perawatanrambut keiskeiindonesia obatkebotakan tokopedia promo diskon amazing	6
...	...	...

Proses selanjutnya setelah data *tweet* dikelompokkan ke dalam 8 *Cluster* adalah menampilkan 10 *top word* pada masing-masing *Cluster*. Berikut merupakan hasil perhitungan *top word* dapat dilihat pada Tabel 4.

Tabel 4. *Top word* pada setiap *Cluster*.

<b>Cluster</b>	<b>Kata</b>
0	shopee, celotehads, freeongkir, celotehshop, varianrasa, 500gr, ready, asli, wajit, cililin
1	qoo10, cracked, docomo, amazon, zagg, bodyguardz, pixel, iphone, galaxy, screen
2	80, 200, howtoperfect, howtobeauty, theordinary, newgfr, shopee, shopeemy, himalayamy, himalaya
3	08983840434, 1thn, shopee, celotehshop, celotehads, freeongkir, garansi, expro, maticbebek, kursibocenganak
4	mega, pink, marlowe, for, rt, coins, blackpinkforshopeeph, shopeemegashoppingday, blackpink, shopee
5	to, in, 279, shopeeshopeepsreview, shopeeth, tokopedia, rt, lazada, bukalapak, shopee
6	deklimisterpomadeshop, cek, order, kota, sekarang, mejaevent, via, lewat, jual, tokopedia
7	netflixalhambra, thevoice, samsung, original, lazadagonghivesalon, hiveprofessional, 555, gonghivesalon, rt, lazada

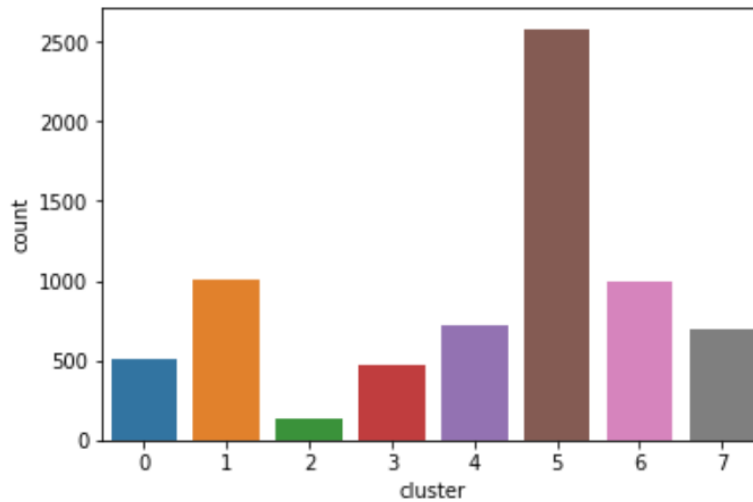
Berdasarkan hasil *top word* yang sudah diperoleh pada Tabel 4., maka dapat ditentukan jenis konten yang terdapat pada setiap *Cluster*. Proses penentuan jenis *Cluster* berdasarkan tingkat kemunculan kata pada setiap *cluster* dapat dilihat pada Tabel 5.

Tabel 5. Tabel konten setiap *Cluster*.

<b>Cluster</b>	<b>Konten</b>
0	Membahas mengenai layanan yang diberikan oleh Celotehsop, serta membahas mengenai produk makanan kas wajit.
1	Membahas produk-produk yang berkaitan dengan <i>smartphone</i> .
2	Membahas mengenai produk-produk kecantikan dan fashion.
3	Membahas produk-produk yang dijual oleh Celotehshop.
4	Membahas mengenai grup music Blackpink dan koin Shopee.
5	Membahas mengenai <i>marketplace</i> yaitu Shopee, Lazada, Tokopedia, dan Bukalapak.
6	Membahas mengenai produk perawatan rambut dan membahas mengenai <i>marketplace</i> Tokopedia.
7	Membahas tentang <i>marketplace</i> Lazada.Lazada

Berikut tampilan visual mengenai hasil *Clustering* menggunakan metode *K-Means* menggunakan diagram batang dapat dilihat pada Gambar 3.





Gambar 3. Diagram Cluster.

Setelah menemukan jenis konten pada setiap *Cluster*, langkah selanjutnya adalah menghitung frekuensi kemunculan kata, dimana tingkat kemunculan kata semakin tinggi maka kata tersebut akan menjadi sebuah tren dalam perbincangan di media sosial Twitter. Berdasarkan hasil perhitungan frekuensi, maka dapat disimpulkan bahwa tren *marketplace* adalah Shopee, Lazada, dan Tokopedia. Berikut merupakan 10 urutan kemunculan kata terbanyak yang diperoleh pada semua *cluster*, dapat dilihat pada Tabel 6.

Tabel 6. Tabel Frekuensi kata.

Kata	Frekuensi
shopee	4890
lazada	2549
tokopedia	2359
rt	2155
galaxy	1471
screen	1375
iphone	1239
pixel	1109
amazon	1061
bodyguardz	1021

#### 4. Kesimpulan

Penelitian ini telah berhasil melakukan analisis tren percakapan market place pada *platform* twitter. Berdasarkan pada hasil eksperimen, diketahui bahwa yang menjadi tren *marketplace* adalah Shopee, kemudian Lazada, dan Tokopedia. Hasil dari perhitungan *tweet* pada setiap *cluster* diperoleh 3 *cluster* dengan *tweet* terbanyak yaitu *Cluster* 5 dengan konten yang membahas mengenai *marketplace* yaitu Shopee, Lazada, Tokopedia dan Bukalapak. *Cluster* 0 dengan konten yang membahas mengenai produk-produk yang berkaitan dengan *smartphone*. *Cluster* 6 dengan konten yang membahas mengenai produk perawatan rambut serta membahas *marketplace* Tokopedia. Dari hasil penelitian ini didapatkan bahwa penerapan metode K-Means

*Clustering* dapat menghasilkan informasi yang cukup sebagai dasar pertimbangan pelaku bisnis dalam memilih *marketplace*.

## Daftar Pustaka

- [1] Sunyoto, D. (2011). *Analisis Regresi dan Uji Hipotesis*. Yogyakarta: Caps Publishing.
- [2] Databoks. (2019). Berapa Pengguna Media Sosial Indonesia? Retrieved November 30, 2019, from <https://databoks.katadata.co.id/datapublish/2019/02/08/berapa-pengguna-media-sosial-indonesia>
- [3] Clinten, B. (2019). Pengguna Aktif Harian Twitter Indonesia Diklaim Terbanyak. Retrieved November 30, 2019, from <https://tekno.kompas.com/read/2019/10/30/16062477/pengguna-aktif-harian-twitter-indonesia-diklaim-terbanyak>
- [4] Habibi, M. (2018). Analisis Konten Jejaring Sosial Twitter dalam Kasus Pemilihan Gubernur DKI 2017. *Teknomatika*, 11(1), 31–40.
- [5] Cahyo, P. W. (2017). *Model Monitoring Sebaran Penyakit Demam Berdarah di Indonesia Berdasarkan Analisis Pesan Twitter*. Universitas Gadjah Mada Yogyakarta.
- [6] Habibi, M., & Cahyo, P. W. (2019). Clustering User Characteristics Based on the influence of Hashtags on the Instagram Platform. *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, 13(4), 399–408. <https://doi.org/10.22146/ijccs.50574>
- [7] Cahyo, P. W. (2018). Klasterisasi Tipe Pembelajar Sebagai Parameter Evaluasi Kualitas Pendidikan di Perguruan Tinggi. *Teknomatika*, 11(1), 49–55.
- [8] Maulida, L. (2018). Kunjungan Wisatawan Ke Objek Wisata Unggulan Di Prov . Dki Jakarta Dengan K-Means. *JISKA*, 2(3), 167–174.
- [9] Kelik Nugroho, A. (2018). Image Quantization in Psoriasis Using K-Mean Clustering. In *Prosiding Seminar Nasional Teknologi Informasi dan Kedirgantaraan (SENATIK)* (Vol. IV, pp. 183–189). <https://doi.org/10.1017/CBO9781107415324.004>
- [10] Habibi, M., & Sumarsono, M. (2018). Implementation of Cosine Similarity in an automatic classifier for comments Program. *JISKA* (Vol. 3). <https://doi.org/http://dx.doi.org/10.14421/jiska.2018.32-05>
- [11] Sebastiani, F. (2002). Machine Learning in Automated Text Categorization. *ACM Computing Surveys (CSUR)*, 34(1), 1–47.
- [12] Habibi, M. (2017). Analisis Sentimen dan Klasifikasi Komentar Mahasiswa pada Sistem Evaluasi Pembelajaran Menggunakan Kombinasi KNN Berbasis Cosine Similarity dan Supervised Model. *Departemen Ilmu Komputer dan Elektronika, Fakultas Matematika dan Ilmu Pengetahuan Alam*. Universitas Gadjah Mada.
- [13] Kalra, V., & Aggarwal, R. (2018). Importance of Text Data Preprocessing & Implementation in RapidMiner. *Proceedings of the First International Conference on Information Technology and Knowledge Management*, 14, 71–75. <https://doi.org/10.15439/2017km46>

- [14] Siqueira, H., & Barros, F. (2010). A Feature Extraction Process for Sentiment Analysis of Opinions on Services. *Proceedings of the III International Workshop on Web and Text Intelligence (WTI)*.
- [15] Manning, C. D., Raghavan, P., & Schütze, H. (2009). *An Introduction to Information Retrieval*. Cambridge, England: Cambridge University Press.  
<https://doi.org/10.1109/LPT.2009.2020494>
- [16] Nisha, N., & Jai Kaur, P. (2015). A Survey of Clustering Techniques and Algorithms. *2nd International Conference on Computing for Sustainable Global Development (INDIACom)*, 3014–3017.
- [17] Putu, N., Merliana, E., & Santoso, A. J. (2015). Analisa Penentuan Jumlah Cluster Terbaik pada Metode K-Means. In *PROSIDING SEMINAR NASIONAL MULTI DISIPLIN ILMU&CALL FOR PAPERS UNISBANK (SENDI\_U)* (pp. 978–979).