

Penerapan *Text Mining* untuk Melakukan *Clustering Data Tweet* Shopee Indonesia

Dwi Smaradahana Indraloka, dan Budi Santosa

Departemen Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Institut Teknologi Sepuluh Nopember (ITS)

e-mail: budi_s@ie.its.ac.id

Abstrak—Pelaku bisnis dapat memanfaatkan *followers* akun Twitter mereka sebagai sarana untuk melakukan *advertising*. Sebanyak 37,5% pengguna Twitter, menemukan pelaku bisnis baru dikarenakan membaca *retweet* dari *tweet* yang dibuat oleh pelaku bisnis tersebut. Dengan mengetahui jenis konten *tweet* yang banyak dilakukan *retweet* oleh *followers*-nya, pelaku bisnis dapat menggunakan jenis konten *tweet* tersebut sebagai sarana untuk melakukan *advertising* kepada pengguna Twitter. Dalam penelitian ini, dilakukan penerapan *Text Mining* untuk melakukan *clustering* dengan metode *K-means* pada data *tweet* Shopee Indonesia untuk mengetahui jenis konten *tweet* yang banyak dilakukan *retweet* oleh *followers* Shopee Indonesia. Hasil yang didapat adalah jenis konten pada klaster yang memiliki jumlah *retweet* yang tinggi diantaranya tentang kuis berhadiah (klaster 4, klaster 20, dan klaster 28), ulang tahun Shopee Indonesia (klaster 11), dan hobi, kuis, dan *lifestyle* (klaster 7), sehingga Shopee Indonesia dapat menggunakan jenis konten *tweet* tersebut sebagai sarana untuk melakukan *advertising* kepada pengguna Twitter.

Kata Kunci—*Advertising, Clustering, K-means, Text Mining, Twitter.*

I. PENDAHULUAN

ADVERTISING telah berkembang menjadi sistem komunikasi yang sangat penting antara pelaku bisnis dan konsumen. Pelaku bisnis sangat mengandalkan *advertising* untuk membantu memasarkan produk atau layanan yang mereka miliki agar dapat dikenal oleh konsumen. Kelebihan tersebut menjadikan *advertising* memiliki peranan utama dalam program pemasaran [1]. Sebagian besar pelaku bisnis saat ini, menggunakan media sosial sebagai sarana untuk melakukan *advertising* [2].

Twitter merupakan salah satu media sosial paling populer di Indonesia [3]. Berdasarkan data statistik dari Statista Inc., jumlah pengguna aktif Twitter di Indonesia pada tahun 2016 sebanyak 24,34 juta orang. Pengguna Twitter dapat membuat sebuah pesan pendek yang disebut dengan *tweet*, dimana melalui *tweet* tersebut, pengguna Twitter dapat saling berhubungan, berbagi pendapat, dan menemukan kabar dari berbagai penjuru dunia. Sebagian besar pengguna Twitter juga memanfaatkan media sosial ini untuk menemukan pelaku bisnis, dimana mereka akan menjadi pengikut (*followers*) dan berinteraksi dengan pelaku bisnis tersebut [4].

Pelaku bisnis dapat memanfaatkan *followers* akun Twitter mereka sebagai sarana untuk melakukan *advertising*. Dengan menggunakan perintah *retweet* pada Twitter, *followers* dapat menyebarkan *tweet* yang dibuat oleh pelaku bisnis. Sebanyak 37,5% pengguna Twitter, menemukan pelaku bisnis baru

dikarenakan membaca *retweet* dari *tweet* yang dibuat oleh pelaku bisnis tersebut. Pelaku bisnis tentunya dituntut untuk mampu memahami jenis konten dalam sebuah *tweet* yang mendapatkan respon positif dari *followers*, karena setiap kata, foto, dan video yang dicantumkan didalam sebuah *tweet* akan memberikan dampak kepada *followers* [5],[6], dan [7].

Pengumpulan data *tweet* dari Twitter dapat dilakukan dengan mengintegrasikan Twitter API dan R *Studio*. Untuk mempermudah mengetahui jenis konten dari sejumlah data *tweet*, maka perlu dilakukan proses *Text Mining* terhadap data *tweet* tersebut dengan menerapkan teknik *clustering* [8]. Pada *Text Mining*, teknik *clustering* digunakan untuk mengelompokkan data tekstual berdasarkan kesamaan konten yang dimiliki ke dalam beberapa klaster, sehingga didalam setiap klaster akan berisi data tekstual dengan konten semirip mungkin [9].

Salah satu metode yang digunakan untuk melakukan *clustering* adalah metode *K-means*. Keuntungan dari menggunakan metode *K-means* untuk melakukan *clustering* adalah tidak memerlukan jumlah iterasi yang banyak untuk mendapatkan hasil *clustering* yang baik, dimana hal ini sangat tepat untuk diterapkan dalam melakukan *clustering* dengan jumlah data yang besar [10]. Penentuan jumlah klaster terbaik dilakukan dengan menggunakan metode *Silhouette coefficient*. Shopee Indonesia merupakan salah satu pelaku bisnis di Indonesia yang bergerak di bidang *e-commerce* dan menggunakan media sosial Twitter sebagai sarana untuk melakukan *advertising*. Terhitung pada tanggal 1 April 2017, Shopee Indonesia memiliki jumlah *followers* Twitter sebanyak 17,5 ribu dan jumlah *tweet* sebanyak 13,2 ribu. Jumlah tersebut masih sedikit apabila dibandingkan dengan pelaku bisnis pesaing dari Shopee Indonesia, diantaranya OLX Indonesia yang memiliki jumlah *followers* Twitter sebanyak 202 ribu dan jumlah *tweet* sebanyak 20,7 ribu, Tokopedia yang memiliki jumlah *followers* Twitter sebanyak 121 ribu dan jumlah *tweet* sebanyak 26,2 ribu, dan Bukalapak yang memiliki jumlah *followers* Twitter sebanyak 113 ribu dan jumlah *tweet* sebanyak 47,4 ribu. Dengan menemukan jenis konten *tweet* yang banyak dilakukan *retweet* oleh *followers* dari Shopee Indonesia, diharapkan semakin banyak pengguna Twitter yang menjadi konsumen dari Shopee Indonesia.

II. DESKRIPSI PERMASALAHAN

Pengolahan dan pengambilan informasi dari sebuah data tekstual merupakan permasalahan yang tidak dapat diselesaikan dengan menggunakan teknik *Data Mining*.

Diperlukan beberapa tahapan proses tambahan agar data tekstual yang memiliki format tidak teratur dapat dilakukan pengolahan dan pengambilan informasi, tahapan proses tersebut dikenal dengan istilah *Text Mining*. Penjelasan tahapan proses *Text Mining* menurut Feldman dan Sanger (2007) adalah sebagai berikut:

A. 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. Proses yang dilakukan pada tahapan ini antara lain:

1. *Case folding* yaitu merubah semua karakter huruf pada sebuah kalimat menjadi huruf kecil dan menghilangkan karakter yang dianggap tidak valid seperti angka, tanda baca, dan *Uniform Resources Locator* (URL).
2. *Tokenizing* yaitu memotong sebuah kalimat berdasarkan tiap kata yang menyusunnya.
3. *Stemming* yaitu merubah berbagai kata berimbuhan menjadi kata dasarnya, tahap ini pada umumnya dilakukan untuk teks dengan bahasa Inggris, karena teks dengan bahasa Inggris memiliki struktur imbuhan yang tetap.
4. *Tagging* yaitu merubah berbagai kata dalam bentuk lampau menjadi kata awalnya, tahap ini pada umumnya dilakukan untuk teks dengan bahasa Inggris atau bahasa lainnya yang memiliki bentuk lampau.

B. Feature selection

Merupakan tahapan untuk mengurangi dimensi dari sebuah data tekstual sehingga hasil dari proses *Text Mining* memiliki kualitas yang lebih baik. Proses yang dilakukan pada tahapan ini adalah *stopword removal* yaitu menghilangkan kata-kata yang dianggap tidak penting atau tidak menggambarkan isi dari sebuah kalimat.

C. Text Representation

Merupakan tahapan merubah data tekstual menjadi representasi yang lebih mudah untuk diproses. Pada tahapan ini, sebuah kalimat direpresentasikan sebagai objek dan kata-kata yang menyusunnya direpresentasikan sebagai fitur. Data tekstual akan membentuk sebuah ruang dengan jumlah objek sebanyak jumlah kalimat yang ada dan jumlah fitur sebanyak jumlah kata yang berbeda.

D. Application of Text Mining Techniques

Merupakan tahapan utama pada proses *Text Mining*. Pada tahapan ini dilakukan penerapan teknik yang digunakan untuk pengambilan informasi dari data tekstual yang telah diproses sebelumnya. Terdapat beberapa teknik yang dapat digunakan diantaranya *classification*, *clustering*, *information extraction*, *trend analysis*, *distribution analysis*, dan *association rules*. Pemilihan teknik yang digunakan disesuaikan dengan jenis informasi yang ingin diambil dari data tekstual yang tersedia.

III. METODE PENYELESAIAN

Algoritma proses *Text Mining* dirancang untuk melakukan *clustering* data *tweet* Shopee Inddonesia. Berikut adalah tahapan proses *Text Mining* yang telah dirancang:

A. Authentication

Pada saat bergabung dengan Twitter API, didapatkan beberapa kode berupa *consumer key*, *consumer secret*, *access token*, dan *access key*.

Tabel 1.
Kode yang didapat pada saat bergabung dengan Twitter API

<i>Consumer key</i>	eqInJRxlWmUCcHFBxWGLT5JUS
<i>Consumer secret</i>	PVIJTrxuIxjAAXIIXYkiEies2vucy1NuUuRnArIDW6c o vb0rYe
<i>Access token</i>	828496585558351872- FPUfdSsBB8dSHLiYoSZqtZXK60v1Q82
<i>Access key</i>	ij6C8mYemRtCUsb6boUBViFBkPFvW0FJtv7q0n9Y g6Rq

Kode di atas digunakan untuk proses integrasi antara Twitter API dengan *software R Studio*, dimana proses integrasi dilakukan dengan menggunakan fungsi ‘*setup_twitter_oauth* (*consumer_key*, *consumer_secret*, *access_token*, *access_key*)’.

B. Pengambilan Data dari Twitter

Pengambilan data dilakukan secara *real time* dari Twitter dengan menggunakan fungsi ‘*tweets = userTimeline* (“ShopeeID”, *n* = 2000, *excludeReplies* = TRUE)’.

Tabel 2.
Contoh data *tweet* yang didapatkan dari *timeline* Twitter Shopee Indonesia

Nomer <i>Tweet</i>	Teks <i>Tweet</i>	Tanggal <i>Tweet</i>	Jumlah <i>Retweet</i>
1	Hey Shopeeholics! Mau tahu siapa saja 3 besar jawara video #ParodiIklanShopee? Cek di IG kami >> instagram.com/shopee_id	01/04/20 17 11:40	0
2	Ayo ikutan Kuis kali ini di IG/FB Shopee ya. 2 Orang yg beruntung akan mendapatkan Brush Set dari Shopee loh!	31/03/20 17 7:37	3

C. Case Folding

Pada proses *case folding*, semua huruf pada tiap *tweet* dirubah menjadi huruf kecil dengan menggunakan fungsi ‘*myCorpus = tm_map* (*myCorpus*, *tolower*)’, karakter yang dianggap tidak valid seperti angka dan tanda baca dihilangkan dengan menggunakan fungsi ‘*myCorpus = tm_map* (*myCorpus*, *content_transformer*(*removeNumPunct*))’, dan *Uniform Resources Locator* (URL) dihilangkan dengan menggunakan fungsi ‘*myCorpus = tm_map* (*myCorpus*, *content_transformer*(*removeURL*))’.

Tabel 3.
Contoh hasil proses *case folding*

Nomer <i>Tweet</i>	Teks <i>Tweet</i> Hasil <i>Case Folding</i>
1	hey shopeeholics mau tahu siapa saja besar jawara video parodiiklanshopee cek di ig kami
2	ayo ikutan kuis kali ini di igfb shopee ya orang yg beruntung akan mendapatkan brush set dari shopee loh

D. Tokenizing

Pada proses *tokenizing*, pemotongan kalimat pada *tweet* berdasarkan tiap kata yang menyusunnya dilakukan dengan menggunakan fungsi ‘*tdm = DocumentTermMatrix (myCorpus)*’.

Tabel 4.
Contoh hasil proses *tokenizing*

Nomer Tweet	Hasil Tokenizing	Nomer Tweet	Hasil Tokenizing
1	hey	2	ayo
	shopeeholics		ikutan
	mau		kuis
	tahu		kali
	siapa		ini
	saja		di
	besar		igfb
	jawara		shopee
	video		ya
	parodiiklanshoopee		orang
	cek		yg
	di		beruntung
	ig		akan
	kami		mendapatkan
			brush
	set		
	dari		
	shopee		
	loh		

E. Stopword Removal

Pada proses *stopword removal*, penghilangan kata-kata yang dianggap tidak penting atau tidak menggambarkan isi dari sebuah *tweet* dilakukan dengan menggunakan fungsi ‘*myCorpus = tm_map (myCorpus, removeWords, custom_stopwords)*’.

Tabel 5.
Contoh hasil proses *stopword removal*

Nomer Tweet	Hasil Stopword Removal	Nomer Tweet	Hasil Stopword Removal
1	hey	2	ayo
	shopeeholics		ikutan
	tahu		kuis
	besar		kali
	jawara		igfb
	video		shopee
	parodiiklanshoopee		ya
	cek		orang
	ig		yg
			beruntung
			mendapatkan
			brush
			set
			shopee
			loh

F. Text Representation

Pada proses *text representation*, perubahan data *tweet* menjadi sebuah matriks dengan baris berupa nomor dari *tweet* dan kolom berupa seluruh kata penyusun data *tweet* dilakukan dengan menggunakan fungsi ‘*matriks_kata = as.matrix(tdm)*’.

Berdasarkan hasil *text representation*, banyaknya seluruh kata yang menyusun 498 *tweet* dari Shopee Indonesia adalah sebanyak 1335 kata. Seluruh kata tersebut menjadi variabel dari tiap *tweet*, dengan komponen dari matriks berupa jumlah dari suatu kata yang ada pada tiap *tweet*.

Tabel 6.

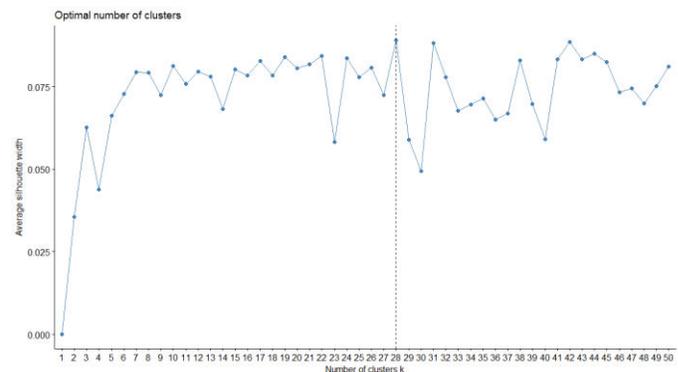
Contoh hasil proses *text representation*

Nomer Tweet	ayo	besar	beruntung	...	video	ya	yg
1	0	1	0	...	1	0	0
2	1	0	1	...	0	1	1

G. Penentuan Jumlah Kluster Terbaik

Penentuan jumlah kluster terbaik dilakukan berdasarkan hasil perhitungan nilai *Silhouette Coefficient* dengan menggunakan fungsi ‘*fviz_nbclust (matriks_kata, kmeans, method = "silhouette", k.max=50)*’.

Perhitungan nilai *Silhouette Coefficient* dilakukan untuk k = 1 sampai k = 50 dan didapat k terbaik adalah k = 28. Hasil perhitungan k terbaik digunakan untuk proses *clustering* dengan *K-means*.



Gambar 1. Nilai *Silhouette Coefficient* pada Jumlah Kluster Sebanyak 1 sampai 50.

H. Clustering dengan K-means

Proses *clustering* dengan *K-means* dilakukan dengan menggunakan fungsi ‘*kmeans_i = kmeans (matriks_kata, 28, 100)*’ dengan jumlah kluster sebanyak 28 kluster dan jumlah iterasi sebanyak 100 iterasi.

Tabel 7.
Hasil *clustering* dengan *k-means*

Nomor Kluster	Nomer Tweet	Jumlah Tweet
1	56, 67, 68, 69, 74, 80, 82, 84, 86, 89, 90, 93,98, 103, 106	15
2	2, 9, 34, 39, 43, 46, 52	7
3	331, 392, 396, 403, 405, 407, 412,413	8
⋮	⋮	⋮
26	11, 61, 232, 279, 322, 365, 375, 385, 453	9
27	127, 206, 218, 251, 288, 293, 309, 310, 323, 357, 359, 366, 388, 442, 474, 476, 484, 489	18
28	32, 197, 431, 444, 454, 463, 471, 481, 493	9

I. Penentuan Jenis Konten Tweet tiap Kluster

Proses pencarian kata yang paling sering muncul pada masing-masing kluster dilakukan dengan menggunakan fungsi ‘*cat(names(sort_kata)[1:10])*’, dimana diambil sebanyak sepuluh kata yang paling sering muncul.

Tabel 8.

Kata paling sering muncul dan jenis konten *tweet* pada tiap kluster

Nomor Kluster	Kata Paling Sering Muncul	Jenis Konten Tweet
---------------	---------------------------	--------------------

1	gunakan, kode, voucher, dapatkan, potongan, untuk, kategori, berlaku, hari, ini ayo, ikutan, kuis, di, fbig, shopee, ya, orang, yang,	Kode voucher untuk potongan harga
2	beruntung tunjukkan, foto, kamu, gaya, belanja, terbaik, hari, ini, dapatkan, iphone	Kuis berhadiah
3	ini, dia, pemenang, seller, of, the, month, bulan, vote, terbanyak	Kontes foto saat berbelanja
⋮	⋮	⋮
26	selamat, untuk, para, pemenang, cepat, kirimkan, data, kamu, kami, tunggu tahu, produk, apa, ini, jawab, di, fbig, shopee, dapatkan, paket	Pengumuman pemenang <i>Seller of The Month</i>
27		Pengumuman pemenang kuis
28		Kuis berhadiah

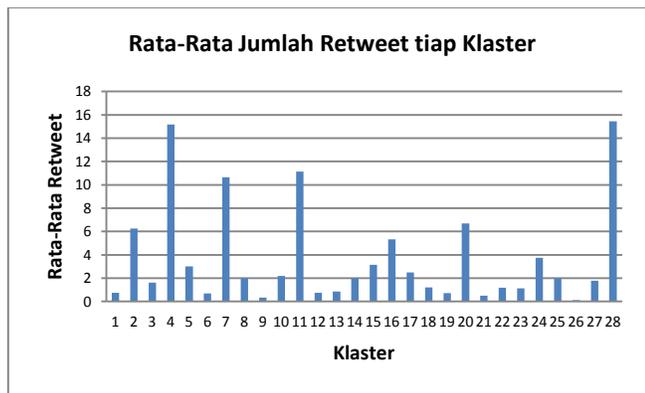
J. Perhitungan Rata-Rata Jumlah Retweet tiap Kluster

Proses perhitungan rata-rata jumlah *retweet* tiap kluster dilakukan terhadap hasil dari *clustering* dengan *K-means*.

Tabel 9. Perhitungan rata-rata jumlah *retweet* tiap kluster

Nomor Kluster	Jumlah Tweet	Total Retweet	Rata-Rata Retweet
1	15	11	0,733
2	7	45	6,249
3	8	13	1,625
⋮	⋮	⋮	⋮
26	9	1	0,111
27	18	32	1,778
28	9	139	15,444

Berikut adalah grafik rata-rata jumlah *retweet* tiap kluster:



Gambar 2. Grafik rata-rata jumlah *retweet* tiap kluster.

Berdasarkan hasil di atas didapatkan bahwa kluster yang memiliki jumlah rata-rata *retweet* tinggi diantaranya kluster 28, kluster 4, kluster 11, kluster 7, dan kluster 20, sedangkan kluster yang memiliki jumlah rata-rata *retweet* rendah diantaranya kluster 26, kluster 9, kluster 21, kluster 1, dan kluster 6. Jenis konten pada kluster yang memiliki jumlah *retweet* yang tinggi diantaranya tentang kuis berhadiah (kluster 4, kluster 20, dan kluster 28), ulang tahun Shopee Indonesia (kluster 11), dan hobi, kuis, dan *lifestyle* (kluster 7), sedangkan jenis konten pada kluster yang memiliki jumlah *retweet* yang rendah diantaranya tentang kode voucher untuk potongan

harga (kluster 1), pengumuman kuis pemenang (kluster 6), *Mobile Shopping Day* (kluster 9), kontes foto *Valentine* (kluster 21), dan pengumuman pemenang *Seller of The Month* (kluster 26).

IV. ANALISIS

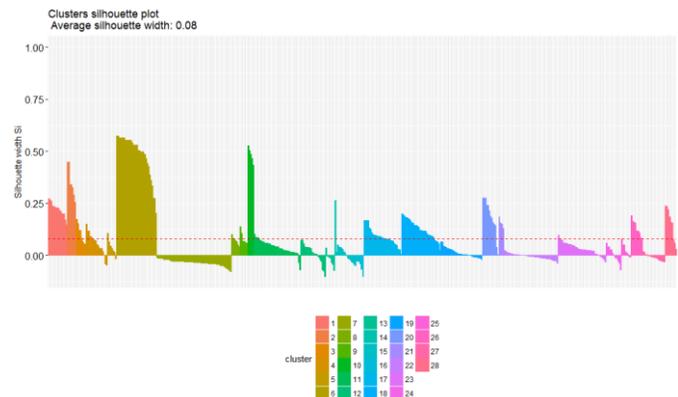
A. Analisis Hasil Clustering dengan Metode K-means

Salah satu metode yang digunakan untuk menguji kualitas kluster yang dihasilkan dari proses *clustering* adalah *Silhouette Coefficient*.

Tabel 10. Perhitungan *silhouette coefficient* pada tiap kluster *tweet*

Nomor Kluster	Jumlah Tweet	Nilai <i>Silhouette Coefficient</i>
1	15	0,22
2	7	0,35
3	8	0,1
⋮	⋮	⋮
26	9	0,13
27	18	-0,01
28	9	0,15

Berikut adalah grafik nilai *Silhouette Coefficient* pada tiap kluster *tweet*:



Gambar 3. Grafik nilai *silhouette coefficient* pada tiap kluster *tweet*.

Berdasarkan hasil perhitungan nilai *Silhouette Coefficient* di atas, diketahui sebanyak 21 kluster memiliki nilai positif, 3 kluster memiliki nilai 0, dan 4 kluster memiliki nilai negatif. Nilai positif menunjukkan bahwa sebagian besar anggota pada kluster berada pada kluster yang tepat, nilai 0 menunjukkan bahwa sebagian besar anggota pada kluster berada di antara 2 kluster, dan nilai negatif menunjukkan bahwa sebagian besar anggota pada kluster berada pada kluster yang tidak tepat. Adanya nilai 0 dan nilai negatif menunjukkan bahwa hasil *clustering* dengan metode *k-means* pada data *tweet* masih belum optimal.

Terdapat beberapa faktor yang membuat hasil *clustering* dengan metode *k-means* pada data *tweet* masih belum optimal, diantaranya adalah sebagai berikut:

1. Batas maksimal karakter yang ada pada sebuah *tweet* adalah 140 karakter, sehingga sebagian besar *tweet*

mengandung kata-kata yang berupa singkatan. Contoh: kata “untuk” menjadi “utk”, kata “yang” menjadi “yg”, kata “dengan” menjadi “dgn” dan lain sebagainya. Penggunaan kata singkatan berpengaruh terhadap proses *stopword removal*, dimana kata yang seharusnya dihilangkan menjadi tidak terdeteksi sehingga akan tetap ada pada kalimat *tweet*.

2. Tidak dilakukannya proses *stemming* dan *tagging* berakibat pada *tweet* yang sebenarnya memiliki maksud yang sama akan dianggap memiliki perbedaan karena penggunaan kalimat aktif dan pasif. Contoh: kalimat “kamu bisa mendapatkan hadiah ini” dan “hadiah ini bisa kamu dapatkan” akan memiliki perbedaan jarak dikarenakan terdapat kata yang berbeda yaitu “mendapatkan” dan “dapatkan”.

Kedua faktor di atas sangat berpengaruh terhadap hasil dari proses *clustering*, dikarenakan setiap kata penyusun dari sebuah *tweet* akan menjadi *variable-variable* yang digunakan pada proses perhitungan jarak antar tiap *tweet*, sehingga hasil *clustering* dapat menempatkan *tweet* pada klaster yang tidak tepat.

B. Analisis Hasil Penentuan Jenis Konten Tweet tiap Klaster

Didapatkan hasil sejumlah 7 klaster memiliki jenis konten *tweet* yang sama yaitu tentang kuis berhadiah. Terdapat beberapa hal yang membuat *tweet* dengan jenis konten yang sama dapat berada pada klaster yang berbeda diantaranya:

1. Perbedaan penggunaan bentuk kalimat aktif dan pasif didalam menyampaikan isi dari *tweet* tersebut.
2. Adanya kata-kata penting yang menjadi pembeda antar tiap klaster seperti nama kuis dan hadiah yang didapatkan.

Meskipun terdapat beberapa klaster dengan jenis konten *tweet* tentang kuis berhadiah, tetapi masing-masing klaster memiliki jumlah *retweet* yang tinggi, sehingga tidak diperlukan analisis lebih lanjut mengenai kuis berhadiah seperti apa yang memiliki jumlah *retweet* tinggi.

C. Analisis Pemilihan Jenis Konten Tweet untuk Sarana Advertising

Berdasarkan hasil perhitungan rata-rata jumlah *retweet* pada tiap klaster, didapatkan bahwa jenis konten pada klaster yang memiliki jumlah *retweet* yang tinggi diantaranya tentang kuis berhadiah, ulang tahun Shopee Indonesia, dan hobi, kuis, dan *lifestyle*, sedangkan jenis konten pada klaster yang memiliki jumlah *retweet* yang rendah diantaranya tentang kode voucher untuk potongan harga, pengumuman kuis pemenang, *Mobile Shopping Day*, kontes foto *Valentine*, dan pengumuman pemenang *Seller of The Month*.

Hasil yang didapatkan menunjukkan bahwa konsumen dari pelaku bisnis Shopee Indonesia lebih tertarik terhadap program kuis berhadiah dibandingkan dengan promo atau acara yang memberikan potongan harga. Sehingga pelaku bisnis Shopee Indonesia diharapkan untuk memanfaatkan jenis konten *tweet* berupa kuis berhadiah sebagai sarana untuk melakukan *advertising* dibandingkan promo atau acara yang memberikan potongan harga. Selain itu, undian atau kuis

berhadiah dapat memberikan dorongan kepada konsumen untuk melakukan pembelian dalam jumlah yang lebih banyak dan meminimumkan perilaku ganti-ganti merek [11]. Sedangkan potongan harga dapat mempengaruhi jumlah penjualan dengan adanya percepatan pembelian, tetapi memiliki dampak negatif terhadap ekuitas merek [12].

V. KESIMPULAN

Kesimpulan yang dapat ditarik pada penelitian ini adalah sebagai berikut:

1. Penerapan algoritma proses *Text Mining* untuk melakukan *clustering* dengan metode *K-means* pada data *tweet* Shopee Indonesia menghasilkan sejumlah 28 klaster *tweet*.
2. Berdasarkan hasil proses penentuan jenis konten dan perhitungan rata-rata jumlah *retweet* pada tiap klaster, didapatkan bahwa jenis konten pada klaster yang memiliki jumlah *retweet* yang tinggi diantaranya tentang kuis berhadiah (klaster 4, klaster 20, dan klaster 28), ulang tahun Shopee Indonesia (klaster 11), dan hobi, kuis, dan *lifestyle* (klaster 7).
3. Pelaku bisnis Shopee Indonesia dapat mengetahui jenis konten *tweet* yang banyak dilakukan *retweet* oleh *followers*-nya sehingga dapat menggunakan jenis konten *tweet* tersebut sebagai sarana untuk melakukan *advertising* kepada pengguna Twitter.

LAMPIRAN

Berikut adalah jenis konten *tweet* pada tiap klaster:

Tabel 11.
Kata paling sering muncul dan jenis konten *tweet* pada tiap klaster

Nomor Klaster	Kata Paling Sering Muncul	Jenis Konten <i>Tweet</i>
1	gunakan, kode, voucher, dapatkan, potongan, untuk, kategori, berlaku, hari, ini ayo, ikutan, kuis, di, fb,ig, shopee, ya, orang, yang, beruntung	Kode voucher untuk potongan harga
2	tunjukkan, foto, kamu, gaya, belanja, terbaik, hari, ini, dapatkan, iphone mau, yuk, ikutan, kuisnyashebi, shopee, keren, kamu, klik, acara, beli	Kuis berhadiah
3	terima, kasih, media, acara, hari, ceo, shopee, indonesia, chris, feng	Kontes foto saat berbelanja
4	selamat, kepada, pemenang, kuis, kali, ini, untuk, konfirmasi, silahkan, kirim shebi, kamu, nonton, film, shopee, ayo, satu, bikin, cari, hobby	Kuis berhadiah
5	shopeeholics, seller, yang, kota, ayo, ikutan, kampus, shopee, roadshow	<i>Press conference</i> CEO Shopee Indonesia
6	di, shopee, mobile, shopping, day, prep, yuk, tips, contest, lihat	Pengumuman pemenang kuis
7	yakin, udah, tahu, shopee, jawab, pertanyaan, ini, di, igfb, dapatkan	Hobi, kuis, dan <i>lifestyle</i>
8		Kampus Shopee <i>Roadshow</i>
9		<i>Mobile Shopping Day</i>
10		Kuis berhadiah

11	conference, shopee, birthday, shopeebirthday, cek, live, facebook, yuk, beli, fbig gunakan, kode, voucher, dptkn, diskon, utk, belanja, untuk, kategori, pria	Ulang tahun Shopee Indonesia	25	lengkapnya ada, unboxing, google, pixel, di, facebook, shopee, live, jam, wib	<i>Unboxing Google Pixel</i>
12	diskon, utk, belanja, untuk, kategori, pria	Kode voucher untuk dapatkan diskon	26	ini, dia, pemenang, seller, of, the, month, bulan, vote, terbanyak	Pengumuman pemenang <i>Seller of The Month</i>
13	shopee, voucher, grosir, mulai, februari	Bulan Cinta Diskon			
14	crazy, gadget, flash, sale, cuma, jam, aja, besok, bakal, acc hari, ada, samsung, galaxy, diskon, harga, pukul, wib, stock, terbatas	<i>Crazy Gadget Flash Sale</i>			
15	ayo, ikutan, kuis, di, fbig, shopee, ya, orang, yang, beruntung	Diskon produk dengan <i>stock</i> terbatas			
16	happy, weekend, shopeeholics, selamat, hari, untuk, merayakan, quoteoftheday, have, day	Kuis berhadiah			
17	mobile, accessories, flash, sale, hari, ini, dimulai, jam, cek, koleksinya	Ucapan selamat <i>happy weekend</i> dan hari spesial			
18	dapatkan, berbagai, shopee, klik, beli, untuk, hari, ini, aja, loh	<i>Mobile Accessories Flash Sale</i>			
19	ayo, ikutan, kuis, di, fbig, shopee, ya, orang, yang, beruntung	Promosi beberapa produk			
20	valentine, sang, buah, hati, ayo, upload, foto, caption, paling, kreatif	Kuis berhadiah			
21	ayo, cobain, bikin, nih, shebi, kasih, lihat, reseponya, untuk, kamu	Kontes foto <i>Valentine</i>			
22	ayo, belanja, koleksi, termurah, shopee, diskon, loh, yuk, cek, koleksinya	Resep masakan dari Shebi			
23	mau, merchandise, film, yuk, cek, fbig, shopee, untuk, info,	Promo diskon beberapa produk			
24		Kuis berhadiah			

DAFTAR PUSTAKA

[1] E. B. George and A. B. Michael, *Advertising and Promotion: An Marketing Communications Perpective*. New York: The McGraw-Hill Companies, 2003.

[2] S. Neti, "Social Media and its Role in Marketing," *J. Enterp. Comput. Bus. Syst.*, vol. 4, 2011.

[3] Kementrian Perdagangan, "Panduan Optimalisasi Media Sosial Untuk Kementrian Perdagangan RI," Jakarta, 2014.

[4] M. Rani and J. Arora, "Twitter Data Predicting Geolocation Using Data Mining Techniques," *Int. J. Innov. Res. Comput. Commun. Eng.*, 2016.

[5] Twitter, "Analitika Twitter," 2016. [Online]. Available: <https://business.twitter.com/id/analytics.html>.

[6] Twitter, "Company," 2016. [Online]. Available: <https://about.twitter.com/id/company>.

[7] Twitter, "Getting Started With Twitter," 2016. [Online]. Available: <https://help.twitter.com/forms>.

[8] Y. Zhao, *R and Data Mining: Examples and Case Studies*. Amsterdam: Elsevier, 2013.

[9] R. Feldman and J. Sanger, *The Text Mining Handbook: Advance Approaches in Analyzing Unstructured Data*. New York: Cambridge University Press, 2007.

[10] C. C. Aggarwal and C. Zhai, *Mining Text Data*. New York: Springer US, 2012.

[11] F. Tjiptono, *Brand Management & Strategy*. Yogyakarta: Andi Offset, 2005.

[12] J. Manel, J., M.C. and S. Zacharias, "Sales Promotion Good or Bad?," in *International Marketing Conference on Marketing & Society*, 2007.