

## **Penerapan *Text Mining* Untuk *Advertising* Pada Data *Tweets* Zalora Indonesia Dengan Menggunakan Metode *K-Means Clustering***

**Feybiola Agustine Andrea Ompo<sup>1\*</sup>, Magdalena A. Ineke Pakereng<sup>2</sup>**  
 Teknik Informatika, Universitas Kristen Satya Wacana, Salatiga, Indonesia  
<sup>\*</sup>e-mail *Corresponding Author*: 672019267@student.uksw.edu

### **Abstract**

*Zalora Indonesia is one of the online retail business people who use Twitter social media as a means to do advertising. The purpose of this study is to determine the type of tweets content that is widely liked and retweeted by Zalora Indonesia followers to advertise to Twitter users. The collection of tweets data is done by integrating the Twitter API and Python programming language. The data analysis method is carried out by utilizing 2 tools, namely the Python programming language for text preprocessing and Rapidminer for data processing using the K-Means algorithm. The results of the application of the K-Means algorithm are 4 clusters, including Zalora cashback (cluster 1), skincare and woman style (cluster 2), payday and shopping time (cluster 3), as well as holiday promos (cluster 4). Based on the calculation of the average number of likes and retweets in each cluster, the type of content with the most likes and retweets was obtained, namely regarding holiday promos (cluster 4). So that business people can take advantage of the like and retweet features as a means for advertising to Zalora Indonesia users.*

**Kata kunci:** *Zalora Indonesia; Text Mining; Clustering; K-Means; Twitter*

### **Abstrak**

Zalora Indonesia merupakan salah satu pelaku bisnis retail *online* yang menggunakan media sosial *Twitter* sebagai sarana untuk melakukan *advertising*. Tujuan dari penelitian ini yaitu untuk mengetahui jenis konten *tweets* yang banyak dilakukan *likes* dan *retweet* oleh *followers* Zalora Indonesia untuk melakukan *advertising* kepada pengguna *Twitter*. Pengumpulan data *tweets* dilakukan dengan mengintegrasikan *Twitter API* dan bahasa pemrograman *Python*. Metode analisis data dilakukan dengan memanfaatkan 2 *tools* yaitu bahasa pemrograman *Python* untuk proses *text preprocessing* serta *Rapidminer* untuk pengolahan data menggunakan algoritma *K-Means*. Hasil dari penerapan algoritma *K-Means* terdapat 4 klaster, diantaranya mengenai *cashback* zalora (*cluster 1*), *skincare* dan *woman style* (*cluster 2*), *payday* dan waktu belanja (*cluster 3*), serta promo hari raya (*cluster 4*). Berdasarkan perhitungan jumlah rata-rata *like* dan *retweet* pada tiap klaster, diperoleh jenis konten dengan *like* dan *retweet* terbanyak yaitu mengenai promo hari raya (*cluster 4*). Sehingga pelaku bisnis dapat memanfaatkan fitur *like* dan *retweet* sebagai sarana untuk *advertising* kepada pengguna Zalora Indonesia.

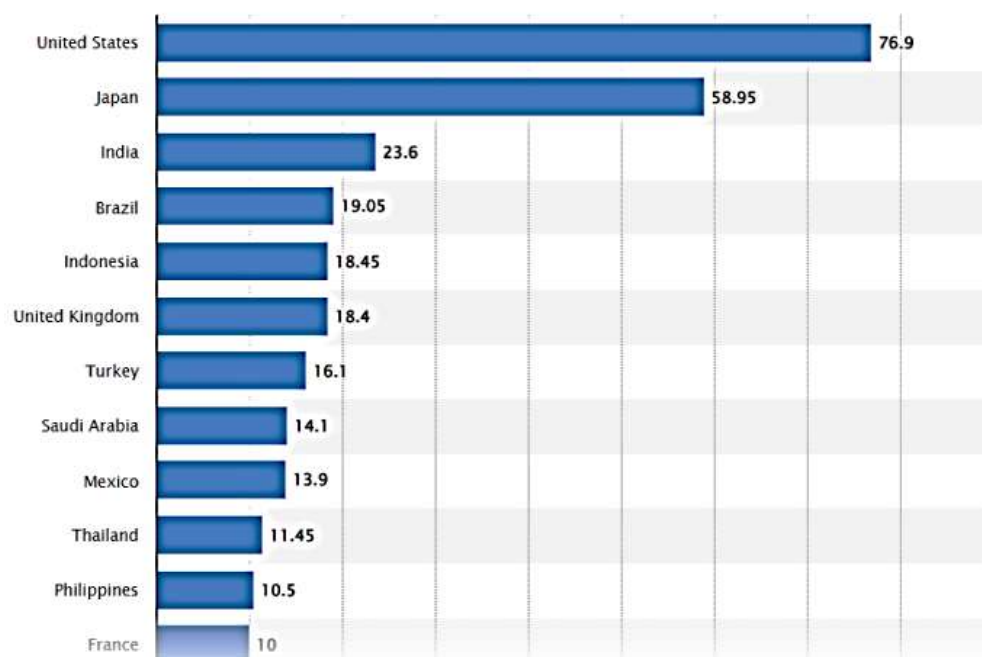
**Kata kunci:** *Zalora Indonesia; Text Mining; Clustering; K-Means; Twitter*

### **1. Pendahuluan**

Saat ini media sosial merupakan salah satu kebutuhan utama yang tidak lepas dari kehidupan sehari-hari. Banyak jejaring sosial yang kini banyak diminati oleh masyarakat termasuk *Twitter*. *Twitter* adalah salah satu jejaring sosial *microblogging* yang memudahkan penggunaannya untuk menulis dan memposting berbagai hal berdasarkan aktivitas dan opini mereka [1]. Pada Gambar 1 oleh Statista, pengguna *Twitter* di Indonesia menempati urutan kelima dengan kurang lebih 18 juta pengguna di seluruh dunia.

Bisnis sangat bergantung pada *Twitter* untuk memasarkan produk atau layanan mereka sehingga konsumen mengenalinya. Salah satu fitur promosi *Twitter* adalah *Tweet Promosi*. *Tweet Promosi* adalah *tweet* biasa yang diiklankan oleh pengiklanan dengan membayar

penempatannya di *Twitter*. Iklan *twitter* mudah dipahami pelanggan yaitu berupa tulisan dan teks dan dapat ditampilkan sampai dengan 140 karakter, melalui iklan *twitter* juga pelanggan dapat langsung merespon iklan yang ditampilkan melalui media sosial *twitter* tersebut [3]. *Twitter* nantinya akan menentukan *Tweet* Promosi mana yang relevan bagi pengguna berdasarkan *tweet* dari pengguna, *following*, *retweet*, pencarian, maupun berbagai interaksi dengan *tweet* atau akun lainnya di *Twitter* [4]. *Twitter* memiliki akses ke *platform* API (Application Programming Interface) untuk menerima data dengan cepat dan dalam jumlah besar untuk tujuan analitik.



Gambar 1. Grafik Pengguna Aktif Twitter Tahun 2022 [2]

Zalora Indonesia merupakan salah satu pelaku bisnis bidang *e-commerce* yang memanfaatkan *Twitter* sebagai sarana untuk melakukan *advertising*. *Username* @zaloraidsn di *Twitter* telah memiliki jumlah *followers* sebanyak 9.5 ribu orang dan jumlah *tweet* sebanyak 3.9 ribu *tweets*. Dengan menemukan berbagai jenis konten *tweet* dari *retweet* yang dilakukan oleh *followers* Zalora Indonesia, diharapkan akan semakin banyak pengguna *Twitter* yang menjadi konsumen dari Zalora Indonesia [5]. Untuk mempermudah mengetahui jenis konten dari sejumlah data *Twitter*, maka diperlukan proses *Text Mining* terhadap data *tweet* tersebut dengan menerapkan teknik *clustering* yang nantinya akan digunakan untuk mengelompokkan data tekstual yang berisi konten semirip mungkin. Dalam penerapan teknik *clustering* ini menggunakan *Rapidminer* sebagai tools tambahan untuk memperoleh hasil yang maksimal [6].

*Text mining* merupakan proses analisis teks yang sumber data biasanya berasal dari sekumpulan dokumen dari waktu ke waktu dengan tujuan mencari kata-kata yang dapat mewakili isi dari dokumen sehingga dapat dilakukan analisa keterhubungan, keterkaitan dan kelas antar dokumen [4]. *Text mining* terdiri dari proses ekstraksi informasi terpola yang berasal dari sejumlah besar sumber data teks, seperti dokumen *Word*, *PDF*, kutipan teks, atau bahkan *SMS* [7].

Tujuan dari penelitian ini adalah untuk mengetahui jenis konten *tweet* yang banyak dilakukan *retweet* dan *favorite* oleh *followers* Zalora Indonesia sehingga dapat digunakan sebagai sarana *advertising* kepada pengguna *Twitter*. Metode analisis *text mining* dengan menggunakan *K-Means*. Pengelompokan data menggunakan algoritma *K-Means* dimana dalam pengimplementasiannya memanfaatkan 2 tools, yaitu bahasa pemrograman *Python* untuk pengambilan data dan *text processing*, lalu *Rapidminer* untuk melakukan pengolahan algoritma *K-Means*.

## 2. Tinjauan Pustaka

Penelitian terdahulu dilakukan oleh Yusril dkk berjudul "Implementasi Text Mining Untuk Advertising Dengan Menggunakan Metode *K-Means Clustering* Pada Data Tweets Gojek Indonesia". Dalam penelitian ini, digunakan sebanyak 84 *tweet* dari akun Gojek Indonesia yang kemudian di hasilkan 2 klaster dengan *retweet* terbanya yaitu terkait program kuis dan perkenalan produk Gojek Indonesia [4].

Pada artikel sejenis yang dilakukan oleh Indraloka dkk [8] mengenai penerapan *text mining* untuk melakukan *clustering* data *tweet Shopee* Indonesia menggunakan metode *k-means* sebagai sarana untuk melakukan *advertising* kepada pengguna twitter memperoleh hasil jenis konten pada klaster yang memiliki jumlah *retweet* yang tinggi diantaranya mengenai kuis hadiah (klaster 4, klaster 20, dan klaster 28), ulang tahun *Shopee* Indonesia (klaster 11), serta hobi, kuis, dan *lifestyle* (klaster 7). Hal yang sama juga dibahas oleh Ferdias dkk mengenai analisa *text mining* menggunakan metode *clustering* data *tweet* tokopedia indonesia untuk mengetahui jenis konten yang banyak di respon positif oleh *followers* dan juga menentukan konten apa yang tepat dikeluarkan agar sesuai dengan keadaan konsumen [9].

Pada Penelitian yang dilakukan oleh Rejito dkk membahas tentang penerapan *teks mining* menggunakan algoritma *k-means clustering* pada 885 data *tweet* Tokopedia didapatkan 48 *cluster* yang kemudian dikelompokkan menjadi 5 kelompok besar. Berdasarkan hasil pengelompokan terlihat bahwa konten yang paling menarik adalah konten hadiah kuis dan konten yang paling tidak menarik adalah konten gaya hidup. Hasil dari penelitian ini digunakan untuk mengetahui jenis konten *tweet* apa saja yang diminati konsumen Tokopedia sebagai sarana promosi [10].

Selain itu penelitian dengan judul "Penerapan *K-Means Cluster* dalam Memilih Strategi Promosi Penerimaan Mahasiswa Baru di Masa Pandemi Covid-19 di Fakultas Ilmu Keperawatan Universitas Muhammadiyah Kalimantan Timur" dilakukan oleh Ulfah dkk sebagai upaya untuk menentukan suatu informasi strategi promosi penerimaan mahasiswa baru yang tepat sasaran dan efektif untuk digunakan. Hasil dari penelitian menghasilkan 3 *cluster*, dengan strategi promosi dominan terdapat pada *cluster* 0 dan media promosi yang tepat yaitu dengan cara Media *Online* dan Informasi Langsung [11]. Kemudian penelitian mengenai penentuan strategi promosi universitas Al Azhar Indonesia menggunakan *k-means clustering* juga dilakukan oleh Arsyad dkk. Analisis *K-Means Clustering* terhadap data indeks prestasi kumulatif dan jumlah penghasilan mahasiswa menghasilkan tiga klaster dengan menjalankan strategi promosi, UAI disarankan untuk memperhatikan karakteristik dari setiap klaster serta memaksimalkan penggunaan *social media* dan *website* dengan menerapkan elemen *7C Framework* [12].

Penelitian terdahulu mengenai analisis dan penerapan algoritma *k-means* dalam strategi promosi kampus Akademi Maritim Suaka Bahari dilakukan oleh Hartati dkk. Penelitian ini dilakukan untuk mengatasi penumpukkan data dalam jumlah yang besar saat proses penerimaan mahasiswa baru dengan menerapkan *data mining* untuk menghasilkan profil yang memiliki kemiripan atribut yang sama. Hasil yang didapatkan yaitu persebaran anggota *cluster* terbanyak ada pada *cluster* 16 berisi anggota *cluster* 115 *items* [13]. Lalu Penelitian sejenis dilakukan oleh Mahmudi dkk dalam penelitiannya pada Universitas Hasyim Asy'ari untuk melakukan analisis dari beberapa cara promosi yang sudah dilakukan sehingga dapat dilihat strategi promosi mana yang lebih tepat dan efektif dengan menggunakan algoritma *k-means clustering*. Pada penelitian ini menghasilkan jumlah 5 *cluster* ( $k=5$ ) yang diharapkan dapat digunakan sebagai rekomendasi untuk menentukan strategi promosi berdasarkan media promosi masing-masing *cluster* yang terbentuk [14].

Selain itu peneliti Sri Ayu juga melakukan penelitian berjudul "Penerapan *K-Means Clustering* Untuk Strategi Promosi Penerimaan Mahasiswa". Penelitian dilakukan di Kampus ITBI Milenial Medan dan serdang bedagai guna memperoleh informasi mengenai perencanaan dan perancangan strategi promosi yang lebih baik dan tepat. Dari penelitian ini diperoleh 3 *cluster* dengan *cluster* 1 sebagai strategi promosi yang paling efektif [15].

Sehingga untuk penelitian ini, peneliti mengusulkan menggunakan metode *K-Means* untuk proses *cluster*, karena *K-Means* terbukti memiliki keuntungan yang lebih dibandingkan metode yang lain yakni tidak memerlukan iterasi yang banyak dalam proses *clustering* untuk mengetahui jenis konten *tweet* yang banyak mendapatkan *like* dan *retweet* oleh *followers* Zalora Indonesia untuk dapat digunakan sebagai sarana *advertising*. Adapun *state of the art* terletak pada penelitian ini terdapat pada perbedaan dataset, jenis parameter, jumlah parameter

input, serta tahapan preprocessing yang dilakukan. *Tools* yang digunakan dari proses pengumpulan data hingga pengolahan data pada artikel ini juga berbeda dengan penelitian [4], yaitu menggunakan 2 *tools* yakni bahasa pemrograman python untuk proses pengumpulan data hingga *preprocessing* dan menggunakan rapidminer untuk proses *clustering*.

### 3. Metodologi

#### 3.1 Jenis dan Sumber Data

Penelitian ini menggunakan metode penelitian kualitatif. Data yang diambil yaitu data *tweet*, tanggal *tweet*, *like*, dan *retweet* dari akun Twitter @zaloraidn. Proses pertama yang dilakukan adalah *authentication* yang dilakukan pada tanggal 22 April 2023 dengan melakukan proses integrasi antara API (*Application Programming Interface*) Twitter dengan bahasa pemrograman Python, sehingga diperlukan API keys dan token dari Twitter terlebih dahulu. Data yang diperoleh merupakan data pada periode 3 Februari 2023 sampai 22 April 2023 dengan jumlah data *tweet* yang diambil adalah 200 tweets.

#### 3.2 Teknik Pengolahan Data

##### 1) *Crawling Data Twitter*

Pengambilan data dilakukan melalui proses *Crawling Data* melalui Twitter API (*Application Programming Interface*) dengan perolehan data 200 *tweets* terbaru di media sosial Twitter @zaloraidn.

##### 2) *Text Pre-Processing*

Tahapan ini merupakan tahapan dimana aplikasi melakukan seleksi dari data teks yang akan diproses. Proses yang dilakukan antara lain:

- *Text Cleaning*

Tahapan membersihkan *tweet* dari *mention*, simbol, url, *hashtag*, angka.

- *Lowering Case*

Tahapan menyeragamkan semua huruf menjadi huruf kecil.

- *Normalization*

Tahapan untuk mengubah kata yang tidak baku menjadi baku.

- *Stopword Removal*

Tahapan menghapus *tweet* yang tidak digunakan dari proses sebelumnya. Cara untuk menghilangkan *tweet* adalah dengan menggunakan kamus kata (*stopword*).

- *Stemming*

Tahapan mengganti semua *tweet* menjadi bentuk kata dasarnya

- *Tokenization*

Tahapan untuk memecah kalimat per kata.

##### 3) *Text Representation*

Pada tahap ini mempresentasikan sebuah kalimat sebagai objek dan kata-kata yang disusun sebagai fitur. Selain itu dilakukan pendekatan model ruang vektor dimana sebuah model mempelajari sebuah kosakata dari seluruh dokumen, lalu akan memodelkan dokumen dengan menghitung jumlah kemunculan dari setiap kata. Pembobotan data teks dilakukan dengan TF-IDF dalam bentuk *term-document matrix*.

##### 4) Menentukan Jumlah Kluster Terbaik

Jumlah kluster terbaik ditentukan menggunakan metode Elbow, dengan melihat persentase hasil yang diperoleh terhadap jumlah kluster yang akhirnya membentuk siku. Pengujian dilakukan dengan jumlah *cluster* uji mulai dari k=2 hingga k=7 dengan menggunakan data dari hasil pembobotan TF-IDF yang menghasilkan matriks kata.

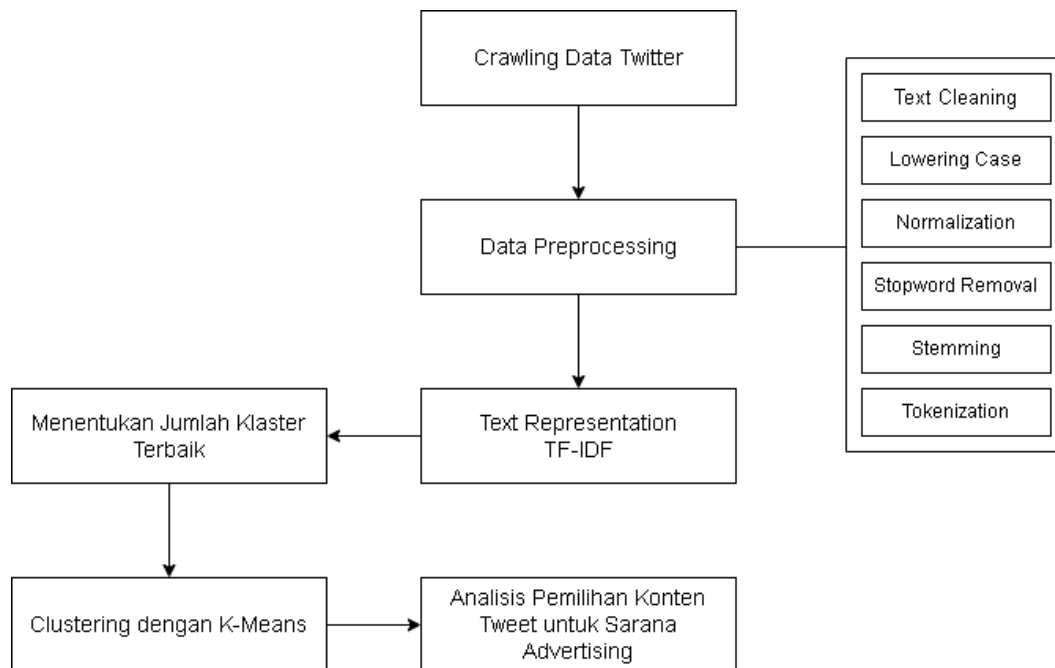
##### 5) *Clustering*

Pada tahap ini dilakukan *clustering* dengan metode *K-Means* dimana metode ini tidak terlalu rumit, bisa mengelompokkan data dalam jumlah yang sangat besar, serta tidak membutuhkan waktu yang lama dalam proses *clustering* objek.

*Clustering* dengan metode *k-means* yaitu dengan mengelompokkan n titik data didalam k *cluster* dengan meminimasi jarak antara titik data dari pusat k *cluster* secara berulang. Langkah-langkah dalam algoritma *k-means clustering* adalah:

1. Menentukan jumlah *cluster*
2. Menentukan nilai *centroid* (pusat *cluster*) Perhitungan jarak dapat dilakukan menggunakan jarak *Euclidean* (Euclidean Distance)
3. Menempatkan titik data pada *cluster* yang memiliki *centroid* dengan jarak minimum
4. Kembali ke tahap 2, lakukan perulangan hingga nilai *centroid* yang dihasilkan tetap dan anggota *cluster* tidak berpindah ke *cluster* lain.

*Flowchart* metode penelitian yang digunakan peneliti digambarkan seperti pada Gambar 2.



Gambar 2. *Flowchart* Metode Penelitian

## 4. Hasil dan Pembahasan

### 4.1 *Crawling* Data Twitter

*Crawling* data Twitter bertujuan untuk mengumpulkan data teks dari aplikasi Twitter dengan memanfaatkan Twitter API. Dalam proses *crawling* data dibutuhkan empat kode akses, yaitu *Consumer Key*, *Consumer Secret*, *Access Token*, dan *Access Token Secret*. Kode yang peneliti peroleh dari Twitter API ditampilkan pada Tabel 1.

Tabel 1. Kode Yang Diperoleh Dari Twitter API

<i>Consumer Key</i>	co9Houpjif2S8B1ROouLQc0X2
<i>Consumer Secret</i>	kJ6RotEyunMJhoUBhrnup7Ue7bKQF0hxGjHETduZzpMjc18iEA
<i>Access Token</i>	1584363252997242881-tptfallZNIYgfEBPoeB0QGwWMMYrR0
<i>Access Token Secret</i>	LeBP0YeAM9XzDKZKmuAg4rFfoz1s5NxCYJZSICBhn6uLW

Kode-kode tersebut diperoleh setelah mendaftarkan akun Twitter pada <https://developer.twitter.com/en/apps>. Kemudian proses integrasi dilakukan menggunakan bahasa pemrograman Python dimana pada prosesnya menggunakan fungsi: 'auth = tweepy.Oauth Handler (consumer\_key, consumer\_secret), auth.access\_token (access\_token, access\_token\_secret)'. Data yang di *crawling* pada akun twitter @zaloraidn dalam periode 3 Februari 2023 sampai 22 April 2023 menghasilkan 200 tweets. Pada Tabel 2 merupakan contoh data *tweets* yang di tampilkan di timeline Twitter Zalora Indonesia.

Tabel 2. Contoh Data Tweets Yang Diperoleh

Nomor Tweet	Teks Tweet	Tanggal Tweet	Jumlah Retweets	Jumlah Likes
19	BELANJA BAJU LEBARAN DADAKAN PAKAI THR DARI ZALORA!GIFTCARD 500K untuk 3 pemenang! 1. Screenshoot dan upload p... <a href="https://t.co/nM18d3LJIS">https://t.co/nM18d3LJIS</a>	2023-04-17 14:38:32+00:00	42	35
20	Siapa yang masih galau belanjain THRnyaa?! 😊 Check out sekarang & serbu Voucher 30% dengan Potongan hingga 500K... <a href="https://t.co/kP0kQcpScg">https://t.co/kP0kQcpScg</a>	2023-04-17 13:00:01+00:00	3	2

#### 4.2 Data Preprocessing

##### 1) Text Cleaning

Pada proses text cleaning dilakukan beberapa fungsi diantaranya seperti pada Tabel 3. Untuk contoh hasil proses text cleaning ditampilkan pada Tabel 4.

Tabel 3. Fungsi Dalam Text Cleaning

No	Fungsi	Syntax
1	Hapus Mention	<code>tweet = re.sub('@[\^s]+', '', tweet)</code>
2	Hapus Simbol	<code>clean_spcl = re.compile('[/(){} @,;:']</code> <code>clean_symbol = re.compile('[^\0-9a-z']</code>
3	Hapus URL	<code>tweet = re.sub('http[\^s]+', '', tweet)</code>
4	Hapus Hastag	<code>tweet = re.sub(r'#([\^s]+)', '', tweet)</code>
5	Hapus Angka	<code>tweet = re.sub(r'[0-9]+', '', tweet)</code>

Tabel 4. Contoh Hasil Proses Text Cleaning

Nomor Tweet	Teks Tweet Sebelum Text Cleaning	Teks Tweet Hasil Text Cleaning
19	BELANJA BAJU LEBARAN DADAKAN PAKAI THR DARI ZALORA!GIFTCARD 500K untuk 3 pemenang! 1. Screenshoot dan upload p... <a href="https://t.co/nM18d3LJIS">https://t.co/nM18d3LJIS</a>	BELANJA BAJU LEBARAN DADAKAN PAKAI THR DARI ZALORAGIFTCARD K untuk pemenang. Screenshoot dan upload p
20	Siapa yang masih galau belanjain THRnyaa?! 😊 Check out sekarang & serbu Voucher 30% dengan Potongan hingga 500K... <a href="https://t.co/kP0kQcpScg">https://t.co/kP0kQcpScg</a>	Siapa yang masih galau belanjain THRnyaa ! Check out sekarang amp; serbu Voucher dengan Potongan hingga K

##### 2) Lowering Case

Pada proses *lowering case* dilakukan fungsi `lwr = lwr.lower()` untuk menyeragamkan semua huruf menjadi huruf kecil seperti yang ditampilkan pada Tabel 5.

Tabel 5. Contoh Hasil Proses *Lowering Case*

Nomor Tweet	Teks Tweet Sebelum <i>Lowering Case</i>	Teks Tweet Hasil <i>Lowering Case</i>
19	BELANJA BAJU LEBARAN DADAKAN PAKAI THR DARI ZALORAGIFTCARD K untuk pemenang. Screenshot dan upload p	belanja baju lebaran dadakan pakai thr dari zaloragiftcard k untuk pemenang screenshot dan upload p
20	Siapa yang masih galau belanjain THRnyaa ! Check out sekarang amp; serbu Voucher dengan Potongan hingga K	siapa yang masih galau belanjain thrnyaa check out sekarang amp; serbu voucher dengan potongan hingga k

### 3) Normalization

Pada proses *normalization* dilakukan *converting chat words* dengan membuat *dictionary* untuk *chat word* beserta kata bakunya untuk menormalisasi kata-kata *chat* yang sering digunakan dalam media sosial seperti btw, otw, fyi, bday dan lainnya. Untuk contoh hasil proses normalisasi ditampilkan pada Tabel 6.

Tabel 6. Contoh Hasil Proses *Normalization*

Nomor Tweet	Teks Tweet Sebelum <i>Normalization</i>	Teks Tweet Hasil <i>Normalization</i>
102	hari terakhir sebelum zalora b day fashion bash berakhir check out sekarang untuk dapatkan extra spesial	hari terakhir sebelum zalora birthday fashion bash berakhir check out sekarang untuk dapatkan extra spesial

### 4) Stopword Removal

Pada proses *stopword removal* dilakukan penghilangan kata-kata dalam korpus yang dianggap tidak menggambarkan isi dari sebuah kalimat. *Stopwords* yang digunakan pada penelitian ini berjumlah 357 kata yang diperoleh <http://www.ranks.nl/stopwords/indonesian>. Serta tambahan *stopwords* sebanyak 75 kata secara manual. Untuk contoh hasil proses *stopword removal* ditampilkan pada Tabel 7.

Tabel 7. Contoh Hasil Proses *Stopword Removal*

Nomor Tweet	Teks Tweet Sebelum <i>Stopword Removal</i>	Teks Tweet Hasil <i>Stopword Removal</i>
9	hi kaka untuk semua produk yang di jual zalora ori ya untuk cara pesannya dapat lihat di	semua produk jual zalora original untuk cara pesannya dapat lihat

### 5) Stemming

Stemming diperlukan untuk menghilangkan imbuhan kata yang ada sehingga dapat memperkecil jumlah indeks yang berbeda dari suatu dokumen. Tabel 8 merupakan contoh proses *stemming* setelah dilakukan *stopword removal*.

Tabel 8. Contoh Hasil Proses *Stemming*

Nomor Tweet	Teks Tweet Sebelum <i>Stemming</i>	Teks Tweet Hasil <i>Stemming</i>
9	semua produk jual zalora original untuk cara pesannya dapat lihat	semua produk jual zalora original untuk cara pesan dapat lihat



**6) Tokenization**

Proses *tokenization* dilakukan untuk memotong tiap kata dari suatu kalimat berdasarkan spasi. Tabel 9 merupakan contoh hasil dari proses *tokenization* setelah dilakukan *stopword removal* dan *stemming*.

Tabel 9. Contoh Hasil Proses *Tokenization*

Nomor Tweet	Teks Tweet Sebelum <i>Tokenization</i>	Teks Tweet Hasil <i>Tokenization</i>
9	semua produk jual zalora original untuk cara pesan dapat lihat	semua produk jual zalora original untuk cara pesan dapat lihat

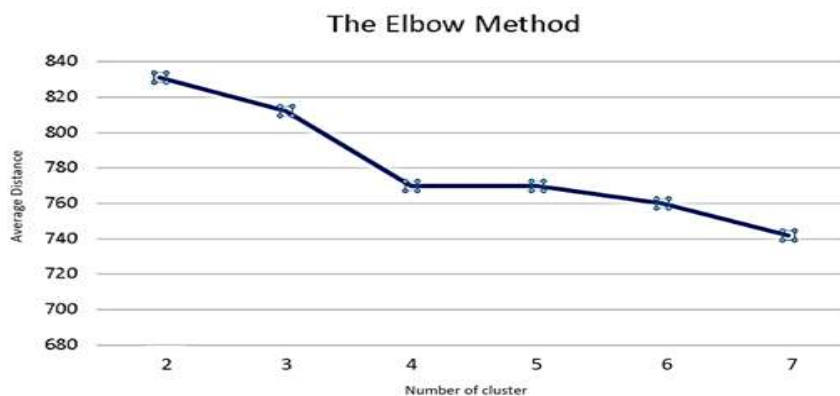
**4.3 Text Representation**

Pada proses *text representation* dilakukan pembobotan kata menggunakan *Term Frequency – Inverse Document Frequency* (TF-IDF). Pembobotan ini nantinya akan digunakan dalam membangun model klasifikasi. Berdasarkan hasil *text representation*, total kata penyusun 200 *tweet* dari Twitter Zalora Indonesia adalah 503 kata. Untuk hasil proses *text representation* ditampilkan pada Tabel 10.

Tabel 10. Contoh Hasil Proses *Text Representation*

	acara	Account	Activated	admin	aesthetic	...	zalora
Tweet ke-1	0	0	0	0	0	...	0.523
Tweet ke-2	0	0	0	0	0	...	0.523
...	...	...	...	...	...	...	...
Tweet Ke-200	0	0	0	0	0	...	0

**4.4 Menentukan Jumlah Kluster Terbaik**



Gambar 3. Metode *Elbow* Pada Jumlah *Cluster* 2 Sampai 7



Metode *Elbow* digunakan untuk menentukan jumlah kluster terbaik pada penelitian ini dengan melihat persentase hasil yang diperoleh dengan membandingkan jumlah kluster pembentuk siku. Pengujian dilakukan dengan jumlah *cluster* uji mulai dari  $k=2$  hingga  $k=7$  dengan menggunakan data dari hasil pembobotan TF-IDF yang menghasilkan matriks kata. Hasil Metode *Elbow* ditampilkan pada Gambar 3.

Pada Gambar 3 terlihat bahwa titik yang paling menyerupai bentuk siku terdapat pada  $k=4$  yang berarti jumlah kluster terbaik adalah 4. Hasil penentuan kluster inilah yang akan digunakan dalam proses *clustering* dengan metode *K-Means*.

**4.5 Clustering dengan K-Means**

Pada proses *clustering* dengan metode *K-Means* data yang diolah merupakan hasil dari proses *preprocessing* dengan jumlah kluster sebanyak 4 dan jumlah iterasi sebanyak 100 iterasi. Untuk hasil *clustering* dengan *K-Means* ditampilkan pada Tabel 11.

Tabel 11. Hasil *Clustering* Dengan *K-Means*

Nomor Kluster	Nomor Tweet	Jumlah Tweets
1	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81	81
2	153, 154, 155, 156, 157, 158, 159, 160, 161, 162, 163, 164, 165, 166, 167, 168, 169, 170, 171, 172, 173, 174, 175, 176, 177, 178, 179, 180, 181, 182, 183, 184, 185, 186, 187, 188, 189, 190, 191, 192, 193, 194, 195, 196, 197, 198, 199, 200	48
3	110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, 123, 124, 125, 126, 127, 128, 129, 130, 131, 132, 133, 134, 135, 136, 137, 138, 139, 140, 141, 142, 143, 144, 145, 146, 147, 148, 149, 150, 151, 152	43
4	82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109	28

Gambar 4, Gambar 5, Gambar 6, dan Gambar 7 menunjukkan penggambaran jenis konten *tweet* pada *cluster 1*, *cluster 2*, *cluster 3*, dan *cluster 4*.



Gambar 4. *Word Cloud Cluster 1*



Gambar 5. *Word Cloud Cluster 2*



Gambar 6. *Word Cloud Cluster 3*



Gambar 7. *Word Cloud Cluster 4*

#### 4.6 Analisis Pemilihan Konten Tweet untuk Sarana Advertising

Penentuan konten yang disukai oleh *followers* akun Twitter @zaloraidn dapat dilihat berdasarkan perhitungan rata-rata dari jumlah *like* dan *retweet* masing-masing *cluster*. Pada Tabel 12 Terlihat perhitungan jumlah *like* dan *retweet* untuk masing-masing *cluster*. Dapat dilihat bahwa jenis konten yang memiliki jumlah rata-rata *like* dan *retweet* tertinggi adalah *cluster* 4 yaitu *tweet* mengenai *wishlist* dan *promo* di bulan *puasa* serta hari raya *nyepi*. Sedangkan *cluster* yang memiliki jumlah rata-rata *like* dan *retweet* terendah adalah *cluster* 3 yaitu mengenai *payday* dan waktu *belanja*.

Tabel 12. Perhitungan Rata-Rata Jumlah Tweet, Like, dan Retweet

Cluster	Jumlah Tweet	Jumlah Like	Jumlah Retweet	Rata-Rata Like	Rata-Rata Retweet	Rata-Rata Total
1	81	81	58	1	0.71	0.85
2	48	41	3	1.46	0.1	0.78
3	43	54	7	1.25	0.16	0.70
4	28	127	17	4.5	0.60	2.55

Hasil yang didapatkan menunjukkan bahwa pelanggan dari Zalora Indonesia lebih tertarik pada konten *promo* hari raya dibandingkan *cashback* zalora, *skincare* & *woman style*, serta *payday* & waktu *belanja*. Hal ini juga dipengaruhi periode waktu data yang diambil yaitu pada bulan Ramadhan dan hari perayaan Nyepi, sehingga ketertarikan pelanggan lebih mengarah ke *promo* untuk produk yang ingin dibeli untuk hari raya yang akan berlangsung. Oleh karena itu untuk memperoleh respon dari *followers* Zalora Indonesia, *tweet* yang dibuat oleh Zalora Indonesia sebaiknya mengenai *promo* untuk produk yang sekiranya akan digunakan pada *event-event* menjelang hari raya atau hari penting lainnya. Dengan membuat *tweet* yang sesuai dengan kebutuhan *followers*, maka akan membantu meningkatkan keinginan *followers* untuk membeli produk dari Zalora Indonesia.

#### 5. Simpulan

Penerapan *algoritma text mining* untuk melakukan *clustering* dengan metode *K-Means* pada data Twitter Zalora Indonesia menghasilkan 4 *cluster* optimal. *Cluster tweet* yang terbentuk adalah konten mengenai *cashback* zalora, *skincare* & *woman style*, *payday* & waktu *belanja*, serta *promo* hari raya. Jumlah anggota dari masing-masing *cluster* tidak jauh berbeda. *Cluster* 1 memiliki anggota sebanyak 81, *cluster* 2 sebanyak 48, *cluster* 3 sebanyak 43, dan *cluster* 4 sebanyak 28.

Berdasarkan perhitungan rata-rata jumlah *likes* dan *retweet* pada tiap klaster yang memiliki jumlah terbanyak adalah *cluster* 4 yaitu mengenai *promo* hari raya untuk memenuhi *wishlist* pelanggan. Hasil perolehan ini tentunya dipengaruhi oleh periode waktu peneliti dalam pengumpulan data. Pelaku bisnis Zalora Indonesia dapat mengetahui jenis konten *tweet* yang banyak mendapatkan *like* dan *retweet* dari *followers*-nya sebagai sarana untuk melakukan *advertising* kepada pengguna Twitter.

Beberapa perbaikan yang dapat dilakukan oleh peneliti selanjutnya adalah pada tahap *preprocessing*. Hal ini dikarenakan tahap *preprocessing* merupakan tahap yang paling penting dan riskan dalam mempengaruhi hasil *clustering*. Perlu dilakukan penambahan kata *slang* yang lebih banyak dalam *dictionary* pada proses *stopword removal* agar hasil *clustering* yang diperoleh jauh lebih akurat.

#### Daftar Referensi

- 1] S. I. Nurhafida and F. Sembiring, "Analisis Text Clustering Masyarakat di Twiter Mengenai Mcdonald'sxbts Menggunakan Orange Data Mining," Semin. Nas. Sist. Inf. dan Manaj. Inform., vol. 1, no. 1, pp. 28–35, 2021.
- [2] J. Clement, "Negara dengan pengguna Twitter terbanyak 2022 | Statistik," Statista, 2022. <https://www.statista.com/statistics/242606/number-of-active-twitter-users-in-selected-countries/>
- [3] S. Dira Febrina Ritonga and S. Wibowo, "Pengaruh Iklan Pada Media Sosial Twitter Dan Televisi Terhadap Keputusan Pembelian Di Mataharimall.com Tahun 2016," e-Proceeding of Applied Science, vol. 2, no. 2, pp. 422–429, 2016.

- 
- [4] A. N. Yusril, I. Larasati, and Q. Aini, "Implementasi Text Mining Untuk Advertising Dengan Menggunakan Metode K-Means Clustering Pada Data Tweets Gojek Indonesia," *Sistemasi*, vol. 9, no. 3, p. 586, 2020, doi: 10.32520/stmsi.v9i3.924.
- [5] S. M. Fani, R. Santoso, and S. Suparti, "Penerapan Text Mining Untuk Melakukan Clustering Data Tweet Akun Blibli Pada Media Sosial Twitter Menggunakan K-Means Clustering," *J. Gaussian*, vol. 10, no. 4, pp. 583–593, 2021, doi: 10.14710/j.gauss.v10i4.30409.
- [6] H. Irsyad and M. R. Pribadi, "Implementasi Text Mining Dalam Pengelompokan Data Tweet Pertanian Indonesia Dengan K-Means," *KURAWAL J. Teknol. Inf. dan Ind.*, vol. 3, no. 2, pp. 164–172, 2020, [Online]. Available: <https://t.co/FXtzMcbdHp>
- [7] N. W. Utami and I. G. J. Eka Putra, "Text Minig Clustering Untuk Pengelompokan Topik Dokumen Penelitian Menggunakan Algoritma K-Means Dengan Cosine Similarity," *Jurnal Informatika Teknologi dan Sains*, vol. 4, no. 3, pp. 255–259, Aug. 2022, doi: 10.51401/jinteks.v4i3.1907.
- [8] D. S. Indraloka and B. Santosa, "Penerapan Text Mining untuk Melakukan Clustering Data Tweet Shopee Indonesia," *Jurnal Sains dan Seni ITS*, vol. 6, no. 2, Sep. 2017, doi: 10.12962/j23373520.v6i2.24419
- [9] P. Ferdias, DE Nurvazly, and SL Chasanah, "Analisa Text Mining Menggunakan Metode Clustering Data Tweet Tokopedia Indonesia," Nov. 2020
- [10] J. Rejito, A. Atthariq, and A. S. Abdullah, "Application of text mining employing k-means algorithms for clustering tweets of Tokopedia," *Journal of Physics: Conference Series*, vol. 1722, no. 1, p. 012019, Jan. 2021, doi: 10.1088/1742-6596/1722/1/012019.
- [11] R. Ulfah, "Penerapan K-Means Cluster dalam Memilih Strategi Promosi Penerimaan Mahasiswa Baru di Masa Pandemi Covid-19 di Fakultas Ilmu Keperawatan Universitas Muhammadiyah Kalimantan Timur," *Fakultas Ilmu Keperawatan Universitas Muhammadiyah Kalimantan Timur*, 2022.
- [12] A. Tiar Arsyad, H. Nurlatifah, and Sunarmo, "Penerapan k-means clustering dalam menentukan Strategi promosi Universitas Al Azhar Indonesia," *Repositori Digital Universitas Al Azhar Indonesia*, 2022.
- [13] T. Hartati, O. Nurdiawan, and E. Wiyandi, "Analisis Dan Penerapan Algoritma K-Means Dalam Strategi Promosi Kampus Akademi Maritim Suaka Bahari," *Jurnal Sains Teknologi Transportasi Maritim*, vol. 3, no. 1, pp. 1–7, May 2021, doi: 10.51578/j.sitektransmar.v3i1.30.
- [14] I. Mahmudi, AD Indriyanti, and I. Lazulfa, "Penerapan Algoritma K-Means Clustering Sebagai Strategi Promosi Penerimaan Mahasiswa Baru Pada Universitas Hasyim Asy'ari Jombang," *Inovate: Jurnal Ilmiah Inovasi Teknologi Informasi*, vol. 4, no. 2, pp. 20–27, April 2020.
- [15] S. Ayu Rosiva Srg, "Penerapan K-Means Clustering Untuk Strategi Promosi Penerimaan Mahasiswa," *JUTISAL Jurnal Teknik Informatika Universal*, vol. 1, no. 1, pp. 1–11, 2021.