

Sentimen Analisis pada Data Tweet Pengguna Twitter Terhadap Produk Penjualan Toko Online Menggunakan Metode K-Means

Andris Faesal¹, Aziz Muslim², Aditya Hastami Ruger³, Kusri⁴

^{1,2,3,4}Program Studi, Megister Teknik Informatika Universitas Amikom Yogyakarta
E-Mail : ¹andris.faesal@students.amikom.ac.id, ²aziz.1122@students.amikom.ac.id,
³aditya.ruger@students.amikom.ac.id, ⁴kusri@amikom.ac.id

ABSTRAK

Pada era big data ini, penggunaan media sosial sering kali membuat postingan di akun media social miliknya berupa opini-opini terhadap kejadian dan barang disekitarnya. Salah satunya yaitu membuat suatu postingan yang memberikan opini pada suatu barang yang baru saja dibeli, sehingga pengaruhnya adalah terhadap pengguna lain yang terhubung dengannya. Semakin banyak orang yang mengetahuinya maka secara tidak langsung orang akan mengenal barang tersebut. Untuk itu, penelitian ini mengangkat sebuah ide yaitu membuat sentiment analisis dari media sosial twitter yang bertujuan untuk memberikan suatu keputusan dari opini konsumen terhadap produk penjualan sehingga menjadi peluang bagi produsen dalam mempromosikan dan memasarkan produknya kepada konsumen. Adapun beberapa tahapan-tahapan dalam penelitian ini dengan menggunakan bahasa pemrograman R dimulai dari pengumpulan data yang dilakukan dengan meng *collect* data tweet pada media social twitter sebanyak 1.000.000 tweet yang berkaitan dengan penjualan di tokopedia, selanjutnya dilakukan pra-prosesing untuk mencari kata-kata yang sering muncul didalam tweet. Langkah terakhir dengan menggunakan metode K-Means sebagai proses pengelompokan sebanyak 3 cluster yaitu kata dengan kemunculan sering, sedang dan jarang digunakan didalam tweet. Dengan pengujian berdasarkan kata tweet diperoleh tingkat akurasi sebesar 92.80 % sedangkan pengujian berdasarkan tweet harian diperoleh tingkat akurasi sebesar 89.80 %.

Kata Kunci : K-Means, Tweet, Tokopedia, Analisis Sentimen.

ABSTRACT

In this big data era, the use of social media often makes posts in his social media accounts in the form of opinions about events and things around him. One of them is making a post that gives an opinion on an item that has just been purchased, so that the effect is on other users who are connected to it. The more people who know it, then indirectly people will get to know the item. For this reason, this study raises an idea that is to make sentiment analysis from Twitter social media that aims to provide a decision of consumer opinion on sales products so that it becomes an opportunity for producers to promote and market their products to consumers. As for some of the stages in this study using the R programming language starting from data collection conducted by collecting tweet data on social media Twitter as many as 1,000,000 tweets related to sales in Tokedia, then pre-processing to search for words which often appears in tweets. The final step is to use the K-Means method as a grouping process of 3 clusters, namely words with frequent, moderate and rarely used in tweets. By testing based on the word tweet obtained an accuracy rate of 92.80% while testing based on daily tweets obtained an accuracy level of 89.80%.

Keyword : K-Means, Tweet, Tokopedia, Sentiment Analysis.

Author Korespondensi (Andris Faesal)
Email : andris.faesal@students.amikom.ac.id

I. PENDAHULUAN

Tokopedia adalah salah satu *market place* yang banyak digunakan oleh masyarakat saat ini. Pada *quarter* pertama tahun 2019 Tokopedia menjadi salah satu dari 10 top brand *e-commerce* yang paling banyak dibicarakan di Twitter. Twitter merupakan media sosial yang sangat populer. Meski begitu Twitter pernah mengalami masa penurunan aktivitas pengguna semenjak munculnya saingan media sosial sekelas Twitter namun memiliki fokus konten gambar yaitu Instagram. Saat ini Twitter mulai banyak digunakan kembali oleh masyarakat. Hal yang paling menarik pada Twitter adalah update nya informasi yang sangat cepat, Hal itu didukung oleh adanya fitur *Trends for you*. Pada fitur ini akan ditampilkan serangkaian kata yang trend dalam arti banyak yang membuat postingan kata tersebut dengan jarak waktu yang kecil sehingga begitu cepat total jumlah posting dengan kata tersebut. Sehingga pengguna Twitter akan mendapatkan informasi yang sesuai dengan trend yang tampil saat itu. Terlepas dari itu semua pengguna juga dapat melakukan pencarian dengan kata untuk sebuah postingan dan atau pengguna lainnya.

Semakin ramai sebuah postingan maka semakin banyak juga pengguna yang terlibat dalam postingan tersebut. Pengguna lain akan dapat dengan mudah memahami topik pembicara dengan beberapa cara menanyakan atau mencari jawaban sendiri dari potongan terkait yang sudah dibuat sebelumnya. Dengan begitu kemudahan sebuah informasi diterima oleh pengguna Twitter melalui trending sangat mudah dilakukan karena kecepatan update yang sangat cepat.

Maka jika topik pembicaraan adalah sebuah produk selanjutnya akan memungkinkan pengguna mencari dimana produk tersebut didapatkan. Sehingga bagi pelaku bisnis yang menggunakan *market place* seperti Tokopedia akan sangat terbantu dalam proses penjualan dan promosi produknya. Namun penulis menemukan masalah yaitu jika pelaku bisnis tersebut salah sasaran ketika mempromosikan produknya dan atau market placency.

Berdasarkan penelitian dari Conor Gallagher dkk pada tahun 2019 tentang sentiment dan text analisis dengan text mining pada data yang berupa *feedback* terhadap product, brand, dan layanannya. Tujuan dari penelitian tersebut

adalah mendapatkan nilai dari pengalaman konsumen dalam merievew barang yang telah dibelinya sehingga didapatkan output berupa strategi bisnis yang dapat meningkatkan nilai kepuasan konsumen dan loyalitas konsumen. Metode yang digunakan adalah naïve Bayes dengan evaluasinya menggunakan confusion matrix dan classification report [1]

Pada topik yang sama yang berhubungan dengan sentimen analisis untuk ulasan produk pada toko online oleh Alvi dkk pada tahun 2017. dalam penelitian tersebut, data yang digunakan adalah review konsumen pada produk amazon untuk kategori telepon selular dan asesorisnya. Proses klasifikasi dengan metode *maximum entropy* dan metode TF-IDF sebagai metode untuk memperoleh fitur pada produk yang telah di review tersebut. Dari hasil pengujian tersebut metode *maximum entropy* dapat mengklasifikasikan opini dari review tersebut menjadi positif dan negatif dengan cukup baik berdasarkan hasil akurasi yang diperoleh sebesar 83% dan untuk *f-1 score* sebesar 90,074%. [2]

Dalam referensi yang lain, penggunaan metode K-Mean sentimen analisis data dari opini 10 destinasi wisata pantai dari twitter oleh Syaifudin dkk pada tahun 2018. dalam penelitian ini, data yang digunakan adalah opini dari pengguna twitter sebanyak 500 tweet yang diklasifikasikan menggunakan algoritma SVM (*Support Vector Machine*) dan menghasilkan nilai akurasi sebesar 74,39% yang selanjutnya dikelompokkan menggunakan metode K-Mean berdasarkan ketersediaan fasilitas yang ada, sumber daya, kesiapan dari masyarakat disekitar, potensi pasar, akses dan posisi pariwisatanya [3]. Hal yang sama juga dibahas oleh Hasyim Siregar ditahun 2018 [4] pada data penjualan alat bangunan.

Sehingga untuk penelitian ini, kami mengusulkan hanya menggunakan metode K-Means untuk proses *cluster* dimana akan ada pra-prosesing terlebih dahulu untuk mengolah data tweet yang bertujuan untuk menghasilkan akurasi yang lebih baik dalam pengelompokan tweet berdasarkan quantity tweet.

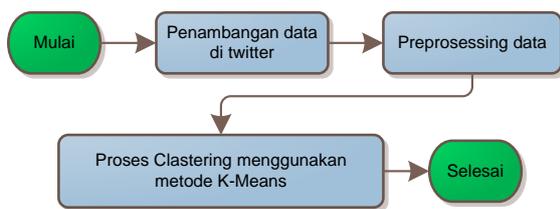
I. METODELOGI

Analisis sentimen merupakan bidang studi yang menganalisa pendapat, sentimen, penilaian, evaluasi, emosi dan sikap seseorang terhadap produk, peristiwa, masalah, topik, individu dan organisasi [5]

Tugas utama dari analisis sentimen yaitu mengelompokkan polaritas atau fitur dalam bentuk text yang terdapat pada kalimat maupun dokumen dan menentukan pendapat yang diuraikan dari kalimat atau dokumen tersebut.

Penelitian ini menggunakan metode penelitian eksperimen, dimana penelitian eksperimen merupakan penelitian yang dilakukan dengan mengadakan manipulasi terhadap objek penelitian serta adanya varibale yang terkontrol [6]. Pada penelitian ini eksperimen yang dilakukan adalah menambang data tweet yang berhubungan dengan penjualan barang tokopedia, selanjutnya data mentah tersebut akan dilakukan proses pra-processing untuk menyaring kata-kata yang tidak penting, menghitung quantity dari masing-masing kata hingga data siap untuk diproses oleh K-means untuk proses pengelompokan data untuk menghasilkan keputusan.

Sistem yang diajukan akan melakukan penambangan data tweet atau *data mining* dari Twitter yang kemudian melalui tahap *preprocessing* sehingga hasil dari *preprocessing* ini akan digunakan untuk proses pengelompokan dengan menggunakan metode K-Means seperti yang ditunjukkan pada gambar 1 berikut



Gambar 1. Alur penelitian

Pada Proses penambangan data twitter untuk memperoleh data tweet yang berhubungan dengan penjualan tokopedia, tools yang digunakan adalah R-Studio dengan bahasa pemrograman R. Untuk dapat mengakses tweet tersebut, dibutuhkan *authentication* twitter (*API-Twitter*) diataranya *consumer_key*, *consumer_secret*, *access_token*, dan *access_secret*.

```
#Memanggil data twett
consumer_key = "xxxxxxxxxxxxxxxxxxxxx"
consumer_secret = "xxxxxxxxxxxxxxxxxxxxx"
access_token = "xxxxxxxxxxxxxxxxxxxxx"
access_token_secret = "xxxxxxxxxxxxx"

#Connect to twitter rest API
setup_twitter_oauth(consumer_key,
consumer_secret, access_token,
access_token_secret)
```

Jumlah tweet yang di panggil untuk di proses sebanyak 1.000.000 (1 Juta) tweet menggunakan kata kunci “tokopedia”. Jumlah tersebut merupakan tweet *random* yang akan di saring sesuai kata kunci yang kita gunakan.

```
tweets = userTimeline('tokopedia', n =
1000000, retryOnRateLimit =
10e6)
```

Dari data Tweet tersebut akan dilakukan *preprocessing* data, yaitu memfilter kata tweet yang tidak sesuai kebutuhan (*anti word*), dalam proses ini akan menghapus kata-kata yang tidak penting yang kemudian digunakan untuk proses selanjutnya.

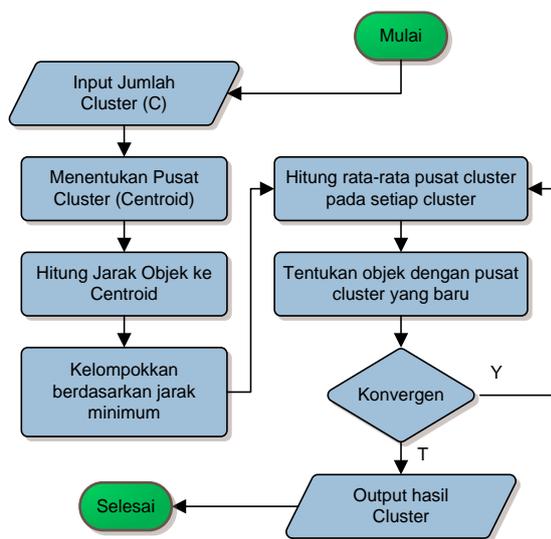
```
myStopwords <-
c(setdiff(stopwords('english'),
c("r", "big")), "kategori", "abis", "ada", "
aja", "apa", "atau", "beberapa", "belum", "b
isa", "blur", "buat", "cuma", "dalam", "dan"
, "dapet", "dari", "dengan", "dem", "dis", "d
pt", "ess", "ini", "itu", "iya", "gak", "jadi"
, "juga", "kamu", "klo", "kalo", "kena", "la
gi", "lain", "lah", "mah", "mau", "nih", "nya"
, "nggak", "org", "orang", "pak", "pake", "p
ara", "perlu", "pas", "penyebab", "pula", "s
ama", "sih", "sini", "sdh", "terus", "trus",
"tuh", "udah", "udh", "utk", "untuk", "yang"
, "yg")
```

Selain proses menghapus kata-kata yang tidak penting pada data tweet, proses yang lain adalah menghilangkan entitas tweet, nama pengguna, tanda baca, angka, link URL, emote dan objek yang tidak penting lainnya.

```
# menyimpan data tweet kedalam variable
myCorpus <-
Corpus(VectorSource(tweets.df$text))
# convert to lower case
myCorpus <- tm_map(myCorpus,
content_transformer(tolower))
# remove URLs
removeURL <- function(x)
gsub("http[^\s:]*", "", x)
myCorpus <- tm_map(myCorpus,
content_transformer(removeURL))
# remove anything other than English
letters or space
```

```
removeNumPunct <- function(x)
gsub("[^[:alpha:][:space:]]*", "", x)
myCorpus <- tm_map(myCorpus,
content_transformer(removeNumPunct))
# remove retweet entities
some_txt =
gsub("(RT|via)((?:\\b\\W*@[\\w+)",
"@\\w+", "[[:punct:]]", "[[:digit:]]",
"http\\w+", "[ \\t]{2,}", "\\s+|\\s+$",
"", some_txt)
```

Untuk langkah terakhir, yaitu menggunakan metode K-Means untuk melakukan proses *clustering*. pengelompokan data berdasarkan jumlah nilai *cluster* yang telah ditentukan, selanjutnya proses konvergen, sehingga ketika data dianggap tidak memenuhi maka dapat dilakukan lagi pemilihan nilai centroid. Tetapi jika data sudah memenuhi maka proses model k-means telah selesai, Berikut adalah Gambar Alur dari Metode K-Means.



Gambar 2. Alur Metode K-Means

Metode ini mempartisi data kedalam kelompok yang lain. Adapun tujuan pengelompokan data ini adalah meminimalkan variasi didalam suatu kelompok dan memaksimalkan variasi antar kelompok. Dengan menggunakan teknik klustering dalam K-means, maka tahapan algoritma K-means adalah sebagai berikut :

- Menentukan koordinat titik tengah setiap cluster
- Menentukan jarak setiap objek terhadap kordinat titik tengah
- Mengelompokan objek-objek tersebut berdasarkan pada jarak minimumnya

- Tentukan pusat cluster baru
- Apakah ada selisih antar cluster lama dengan baru, jika masih ada kembali ke langkah a hingga d, jika tidak selesai.

Di dalam menentukan titik centroid kita menggunakan perhitungan jarak Euclidian distance, dengan rumus sebagai berikut :

$$d_{ij} = \sqrt{[(x_i - x_j)^2 + (y_i - y_j)^2]} \quad (1)$$

Dimana :

- x_i = Koodinat x untuk fasilitas i
- y_i = Koodinat y untuk fasilitas i
- d_{ij} = Jarak antar fasilitas i dan j

Sedangkan untuk mrnghitung nilai centroid cluster dari i dan seterusnya menggunakan rumus berikut :

$$V_{ij} = \frac{\sum_{k=1}^{N_i} X_{kj}}{N_i} \quad (2)$$

Dimana : N_i adalah Jumlah data yang merupakan anggota dari Cluster ke-i

III. HASIL DAN PEMBAHASAN

3.1. Penambangan Data Tweet

Proses penambangan data twitter dilakukan pada tanggal 14 desember 2019, Dari pemanggilan data twitter sebanyak 1 Juta data, (n = 1.000.000), data tweet yang berkaitan dengan kata kunci “tokopedia” berhasil di kumpulkan sebanyak 1.130 tweet yang merupakan konsumen atau akun twitter yang berasal dari Indonesia. Berikut adalah proses *Authentication* twitter dan pemanggilan tweet.

```
1 # step 1
2 # required packages
3 library(twitter)
4 library(sentimentr)
5 library(plyr)
6 library(ggplot2)
7 library(wordcloud)
8 library(RColorBrewer)
9
10 #Authentication twitter (API-Twitter)
11 consumer_key = "RPB3n19L5s0wPINEj05zH35w"
12 consumer_secret = "bVh0w07F3030dwbL3T5mMEar0005ov755vGwK2Tnc80"
13 access_token = "78940374-r5933b1Fm8A8u5lyeSAP081j50xhgV1qCN"
14 access_token_secret = "dVGGQBWkT17e3k82qCQxYoa1vx2MLXopDRmc4032L8"
15
16 setup_twitter_oauth(consumer_key, consumer_secret, access_token, access_token_secret)
17
18
19 # step 2
20 #Memanggil data tweet
21 some_tweets = searchTwitter('tokopedia', n = 1000000, retryOnRateLimit = 10e6)
22 # get the text
23 some_txt = sapply(some_tweets, function(x) x$getText())
24
25
```

Gambar 3. Authentication twitter dan Proses pemanggilan tweet

Berikut adalah beberapa contoh Tweet yang dihasilkan dari proses pada gambar 3 diatas.

sTable 1. contoh hasil data tweet.

No	text	FC	replyToSN	created
1	Selain itu, di #TokopediaBFF juga ada Flash Sale yang bisa kamu ikutin dari jam 10 pagi tadi hingga jam 2 siang nan... https://t.co/vTJA5Tf2sn	6	NA	14/03/2019 13.10
2	Hi Toppers, Tokopedia bekerja sama dengan Top Donasi 200 memungkinkan kalian berpartisipasi untuk mewujudkan mimpi a... https://t.co/vHXjDSMGnn	0	Smart Sakti	09/03/2019 02.15
...
1129	Ayo mana tweet kamu soal #SeberapaMantapKamu ada voucher belanja. Oke mimin mulai ya!	7	NA	09/03/2019 02.01
1130	Buyer: Gimana nih.. pesennya brown malah dikirimnya coklat, lain kali cek dulu sebelum kirim ya.	11	NA	09/03/2019 01.57

3.2. Proses Preprocessing

Pada proses *preprocessing* ini, ada beberapa proses yang dikerjakan, yang pertama yaitu menghilangkan entitas tweet, nama pengguna, tanda baca, angka, link URL, emoticon dan objek yang tidak penting lainnya dan proses kedua yaitu memfilter kata tweet yang tidak sesuai kebutuhan (*anti word*).

Untuk proses pertama, data tweet yang terkumpul akan di hilangkan entitas tweetnya, nama pengguna, tanda baca, angka, link URL, emoticon dan objek yang tidak penting lainnya, gambar berikut adalah proses dari *preprocessing* yang dilakukan

```

25 #step 3
26 # menyimpan data tweet kedalam variable
27 myCorpus <- Corpus(VectorSource(tweets.df$text))
28 # convert to lower case
29 myCorpus <- tm_map(myCorpus, content_transformer(tolower))
30 # remove URLs
31 removeURL <- function(x) gsub("http[[:space:]]*", "", x)
32 myCorpus <- tm_map(myCorpus, content_transformer(removeURL))
33 # remove anything other than English letters or space
34 removeNonPunct <- function(x) gsub("[^[:alpha:][:space:]]*", "", x)
35 myCorpus <- tm_map(myCorpus, content_transformer(removeNonPunct))
36 # remove retweet entities
37 some_txt <- gsub("(RT|via)((?:\b\w*\b)+)", "", some_txt)
38 # remove @ people
39 some_txt <- gsub("@\w+", "", some_txt)
40 # remove punctuation
41 some_txt <- gsub("[[:punct:]]", "", some_txt)
42 # remove numbers
43 some_txt <- gsub("[[:digit:]]", "", some_txt)
44 # remove html links
45 some_txt <- gsub("http\w+", "", some_txt)
46 # remove unnecessary spaces
47 some_txt <- gsub("[[:space:]]+", "", some_txt)
48 # remove special characters
49 some_txt <- gsub("[^a-zA-Z]", "", some_txt)
50 # define "tolower error handling" function
51 try.error <- function(x)
52 {
53   # create missing value
54   y = NA
55   # trycatch error
56   try_error = tryCatch(tolower(x), error=function(e) e)
57   # if no error
58   if (!inherits(try_error, "error"))
59     y = tolower(x)
60   # result
61   return(y)
62 }
63 # lower case using try_error with apply
64 some_txt = apply(some_txt, try_error)
65 # remove NAs in some_txt
66 some_txt = some_txt[!is.na(some_txt)]
67 names(some_txt) = NULL
68 #-----
69

```

Gambar 4. menghapus kata-kata yang tidak penting pada data tweet.

Dari hasil *preprocessing* tersebut, tweet-tweet yang mengandung kata-kata yang tidak penting akan dihapus sehingga data tweet akan berubah menjadi seperti tabel 2 berikut ini.

Tabel 2. Hasil tweet setelah proses *preprocessing* pertama.

No	text	FC	Reply To SN	created
1	selainituditokopediabffjugaa daflashsaleyangbisakamuiku tindarijam10pagitadihinggaj am2siangnan	6	NA	14/03/2019 13.10
2	hitopperstokopediabekerjasa madengantopdonasi200mem ungkinkankalianberpartisipa siuntukmewujudkanmimpia	0	Smart Sakti	09/03/2019 02.15
...
1129	ayomanatweetkamusoalseber apamantapkamuadavoucher elanjaokemiminmulaiya	7	NA	09/03/2019 02.01
1130	buyergimananihpesennya bro wnmalahdikirimnyacoklat lainkalicekdulusebelumkirimya	11	NA	09/03/2019 01.57

Pada proses ke dua yaitu *anti word*, penelitian ini membatasi hanya 14 kata yaitu : "tokopediabt", "tokopediasaja", "mulaijadulu", "toppers", "belanja", "cashback", "instagram", "produk", "ongkir", "gratisongkir", "giveaway", "voucher", "gajian", dan "fashion". Tweet akan di pasring berdasarkan kata-kata tersebut dan akan di hitung frekuensi kemunculan kata-katanya. Berikut adalah hasil proses *preprocessing* ke dua.

Tabel 3. Hasil Parsing dan Jumlah frekuensi kemunculan kata.

No	Word	Jumlah Temuan
1	Tokopediaxpbs	67
2	tokopediasaja	84
3	mulaiajadulu	124
4	Toppers	143
5	belanja	66
6	cashback	63
7	instagram	42
8	produk	54
9	ongkir	29
10	gratisongkir	23
11	giveaway	47
12	voucher	29
13	gajian	13
14	fashion	25

3.2. Proses Clustering dengan K-Means

Pada proses ini, data yang di olah adalah data hasil dari *preprocessing*, dengan menggunakan metode K-Means sesuai alur proses pada gambar 2 dan dengan Jumlah cluster ditentukan sebanyak 3 cluster, maka hasil clustering adalah sebagai berikut.

Kata	Jumlah	Jarak C1	Jarak C2	Jarak C3	Jarak Terpendek	Kelompok
tokopediaxpbs	67	54	9	76	pada C2 (9)	Sedang
tokopediasaja	84	71	26	59	pada C2 (26)	Sedang
mulaiajadulu	124	113	46	19	pada C3 (19)	Sering
toppers	143	130	85	0	pada C3 (0)	Sering
belanja	66	53	8	77	pada C2 (8)	Sedang
cashback	63	50	5	80	pada C2 (5)	Sedang
instagram	42	29	16	101	pada C2 (16)	Sedang
produk	54	42	4	89	pada C2 (4)	Sedang
ongkir	29	16	29	114	pada C1 (16)	Jarang
gratisongkir	23	10	35	120	pada C1 (10)	Jarang
giveaway	47	34	11	96	pada C2 (11)	Sedang
voucher	29	16	29	114	pada C1 (16)	Jarang
gajian	13	0	45	110	pada C1 (0)	Jarang
fashion	25	12	33	118	pada C1 (12)	Jarang

Gambar 5. hasil Clustering dengan K-Means.

Dengan menggunakan seluruh data tweet yang ada, hasil clustering dengan jumlah cluster sebanyak 3 cluster diperoleh bahwa kata “mulaiajadulu” dan “toppers” merupakan kata dengan tweet terbanyak atau sering muncul, disusul untuk kata kunci “tokopediaxpbs”, “tokopediasaja”, “belanja”, “cashback”, “instagram”, “produk”, dan “giveaway” merupakan kata dengan tweet sedang, sedangkan “ongkir”, “gratisongkir”, “voucher”, “gajian”, dan “fashion” merupakan tweet yang frekuensi kemunculannya jarang. Untuk tingkat akurasi sebesar 92.86%.

Pengujian selanjutnya dengan mengelompokkan data tweet berdasarkan hari,

dengan tujuan untuk mengetahui kata tweet mana yang paling sering muncul pada hari tertentu. Dengan proses yang sama pada pengujian sebelumnya diperoleh hasil cluster pada Gambar 6.

Dari hasil pengujian tersebut, dapat diketahui untuk tweet yang sering digunakan hampir di semua hari adalah kata “mulaiajadulu” dan “toppers” kecuali pada hari sabtu hanya kata “toppers” saja yang frekuensi keunculannya sering. Sedangkan untuk kata tweet yang lain tidak jauh berbeda hasilnya dengan pengujian sebelumnya dengan tingkat akurasi sebesar 89.80 %.

Kata	Minggu	Senin	Selasa	Rabu	Kamis	Jumat	Sabtu
tokopediaxpbs	Sedang	Sedang	Sedang	Sedang	Sering	Sedang	Jarang
tokopediasaja	Sedang	Sedang	Sedang	Jarang	Sedang	Sedang	Sedang
mulaiajadulu	Sering	Sering	Sering	Sering	Sedang	Sering	Sedang
toppers	Sering						
belanja	Sedang						
cashback	Sedang						
instagram	Sedang						
produk	Sedang	Jarang	Sering	Sedang	Sedang	Jarang	Jarang
ongkir	Jarang	Jarang	Jarang	Jarang	Sedang	Sedang	Jarang
gratisongkir	Jarang	Jarang	Jarang	Jarang	Sedang	Sedang	Jarang
giveaway	Sedang	Sedang	Jarang	Sedang	Jarang	Sedang	Sedang
voucher	Sedang	Jarang	Jarang	Jarang	Sedang	Sedang	Sedang
gajian	Jarang	Jarang	Sedang	Jarang	Jarang	Jarang	Jarang
fashion	Jarang	Sedang	Jarang	Sedang	Jarang	Sedang	Jarang

Gambar 6. hasil Clustering berdasarkan hari dengan K-Means

Hasil analisa:

1. Clustering pada data tweet untuk kata kunci tokopedia dengan mengelompokkannya menjadi 3 cluster yaitu sering sedang dan jarang. Hal ini memberi arti bahwa kata-kata yang sering muncul merupakan kata yang di sarankan untuk digunakan dalam tweet sebagai media promosi.
2. Clustering dengan kata kunci “mulaiajadulu” dan “toppers” merupakan kata tweet dengan frekuensi kemunculan terbanyak, sedangkan “gajian” dan “gratisongkir” merupakan kata tweet yang paling jarang digunakan dengan tingkat akurasi sebesar 92.86 % dan 89.80 % untuk cluster hariannya.

IV. SIMPULAN DAN SARAN

4.1. Kesimpulan

Berdasarkan hasil pengujian diatas, diperoleh kesimpulan sebagai berikut :

- a. Berdasarkan hasil analisa dari 1.130 tweet yang berhubungan dengan Tokopedia dengan menggunakan metode K-Means untuk 3 kelompok kata berdasarkan frekuensi kemunculannya yaitu sering digunakan, sedang dan jarang digunakan dengan tingkat akurasi sebesar 92.86 %.
- b. Dengan menggunakan proses yang sama pada data tweet harian, diperoleh kata tweet “mulaiajadulu” dan “toppers” merupakan kata yang sering digunakan di hampir semua hari dengan tingkat akurasi sebesar 89.80 %.

4.2.Saran

Penelitian ini masih sangat sederhana, ada beberapa metode clustering yang dapat digunakan pada penelitian ini, selain itu bisa menggabungkan dengan metode yang lain seperti Hirarki Agglomeratif, ANP atau yang lainnya dengan tujuan memperoleh nilai akurasi yang lebih tinggi atau keluaran informasi yang lebih beragam. Terimakasih

REFERENSI

- [1] C. Gallagher, E. Furey, and K. Curran, “The application of sentiment analysis and text analytics to customer experience reviews to understand what customers are really saying,” *Int. J. Data Warehous. Min.*, vol. 15, no. 4, pp. 21–47, 2019, doi: 10.4018/IJDWM.2019100102.
- [2] S. A. F. Alvi Pranandha Syah, Adiwijaya, “Analisis Sentimen Pada Data Ulasan Produk Toko Online Dengan Metode Maximum Entropy Sentiment Analysis on Online Store Product Reviews With Maximum,” *e-Proceeding Eng.*, vol. 4, no. 3, pp. 4632–4640, 2017.
- [3] Y. W. Syaifudin and R. A. Irawan, “Implementasi Analisis Clustering Dan Sentimen Data Twitter Pada Opini Wisata Pantai Menggunakan Metode K-Means,” *J. Inform. Polinema*, vol. 4, no. 3, p. 189, 2018, doi: 10.33795/jip.v4i3.205.
- [4] M. H. Siregar, “Data Mining Klasterisasi Penjualan Alat-Alat Bangunan Menggunakan Metode K-Means (Studi Kasus Di Toko Adi Bangunan),” *Teknol. Dan Open Source*, vol. 1, no. 2, pp. 83–91, 2018, doi: 10.36378/jtos.v1i2.24.
- [5] G. M. J. E. A. F. T. H. R. N. D. Delen, *Practical Text Mining and Statistical Analysis for Non-structured Text Data Applications*, 1st ed. American: Academic Press, 2012.
- [6] J. W. Creswell, *Education Research. Planning, conduction and evaluating quantitative and qualitative research*, 4th ed. Boston: Pearson, 2012.