

ANALISIS TEXT CLUSTERING AKUN FANPAGE SHOPEE INDONESIA DENGAN KOMENTAR FOLLOWERS MENGGUNAKAN TOOLS ORANGE DATA MINING

Ayuni Sentiya¹, Heri Suroyo²

Fakultas Ilmu Komputer, Universitas Bina Darma

Email: ayunisentiya@gmail.com¹, herisuroyo@binadarma.ac.id²

ABSTRAK

Hasil penelitian ini ditemukan fakta bahwa *followers* akun *twitter* Shopee Indonesia paling banyak merespon status yang berhubungan dengan kata kuis hadiah, *retweet* yang memberikan *handphone*, dan tebakan dalam bentuk *vote* dimana topik pembicaraan dalam sebuah *tweet* yang beragam membuat perhatian untuk diteliti. *Text mining* yang digunakan untuk ekstrak secara otomatis dari sumber data teks yang berbeda yang menangani masalah *clustering*. Dengan menggunakan metode *K-Means Clustering* sebagai metode pengelompokan data yang mempartisi data ke dalam bentuk dua atau lebih kelompok. Penelitian ini mencoba menggunakan *text mining* dalam proses untuk pengelolaan serta peringkasan teks sehingga menghasilkan analisis *text clustering* terhadap akun *fanpage* Shopee Indonesia dengan menggunakan *tools Orange Data Mining* dengan memanfaatkan *Preprocess Text* yang meliputi *transformation*, *tokenization*, *normalization*, dan *filtering* yang bertujuan agar teks dapat dibaca dan dianalisis dari *Orange Data Mining* sehingga menghasilkan topik pembahasan dari kata yang dominan muncul dari status dan komentar *followers* Shopee Indonesia. Kumpulan teks dalam jumlah banyak akan di *scraping* dengan menggunakan *Web Scraper*.

Kata kunci: *twitter*, analisis *text clustering*, *orange data mining*, *web scraper*, *shopee indonesia*.

ABSTRACT

The results of this study found the fact that Shopee Indonesia Twitter account followers most responded to the status associated with the word prize quiz, retweet that gave a cellphone, and guesses in the form of votes where the topic of conversation in a diverse tweet made attention to be researched. Text mining is used to extract automatically from different text data sources that handle clustering problems. By using the K-Means Clustering method as a method of grouping data that partitioned data into two or more groups. This research tries to use text mining in the process of managing and summarizing text so as to produce an analysis of text clustering on Shopee Indonesia's fanpage account using Orange Data Mining tools by utilizing Text Preprocess which includes transformation, tokenization, normalization, and filtering which aims to make the text readable and analyzed from Orange Data Mining to produce discussion topics from dominant words that emerge from Shopee Indonesia's followers' status and comments. A large collection of text will be scraped using a Web Scraper.

Keywords: *twitter*, *text clustering analysis*, *orange data mining*, *web scraper*, *shopee indonesia*.

1. PENDAHULUAN

Perkembangan teknologi komunikasi dan informasi di dunia digital yang begitu pesat dirasakan saat ini, merupakan indikasi dari kebutuhan manusia sebagai makhluk sosial yang menginginkan cara efektif dan efisien dalam berkomunikasi dan mendapatkan informasi. Salah satu wadah komunikasi dan informasi dalam dunia digital antara lain adalah jejaring sosial.

Salah satu layanan jejaring sosial yang memiliki fitur unik dalam layanan komunikasi dan informasi adalah layanan jejaring sosial *twitter*. *Twitter* merupakan suatu wadah berkomunikasi dan berbagi informasi, dimana bentuk komunikasi dan informasi dapat disampaikan dalam sebuah *tweet* [6][7]. Banyaknya karakter dalam sebuah *tweet* dibatasi sebanyak seratus empat puluh karakter, oleh karenanya pengguna layanan jejaring sosial *twitter* dituntut untuk menggunakan kata-kata yang singkat, padat dan jelas dalam berkomunikasi dan berbagi informasi dengan sesama pengguna [4].

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. Sebagaimana besar pengguna *twitter* juga memanfaatkan media sosial ini untuk menemukan pelaku bisnis, dimana mereka akan menjadi pengikut (*followers*) dan berinteraksi dengan pelaku bisnis tersebut [8][9].

Shopee Indonesia merupakan salah satu pelaku bisnis di Indonesia yang mengusung model bisnis *marketplace* dengan menggunakan media sosial *twitter* sebagai sarana untuk melakukan promosi terhadap bisnisnya. Dengan menemukan jenis konten *tweet* yang banyak dilakukan *retweet* oleh *followers* sehingga Shopee Indonesia dapat menggunakan jenis konten *tweet* tersebut sebagai saran untuk melakukan promosi kepada pengguna *twitter* serta dapat dengan mudah mengetahui informasi dan membagikan saran maupun kritikan untuk membuat *shopee* menjadi lebih baik lagi dalam proses bisnisnya sehingga diharapkan semakin banyak pengguna *twitter* yang menjadi konsumen dari Shopee Indonesia [2].

Pengumpulan data *tweet* dari *twitter* dapat dilakukan dengan mengintegrasikan *Web Scraping* dan *tools Orange Data Mining*. Untuk mempermudah mengetahui jenis konten dari sejumlah data *tweet*, maka perlu dilakukan proses *text mining* terhadap data *tweet* tersebut dengan menerapkan teknik *clustering* [9]. Pada *text mining*, teknik *clustering* digunakan untuk mengelompokkan data tekstual berdasarkan kesamaan konten yang dimiliki ke dalam beberapa *cluster*, sehingga didalam setiap *cluster* akan berisi data tekstual dengan konten yang mirip.

Berdasarkan uraian diatas maka dibutuhkan analisis informasi pengelompokkan kata yang dominan (sering muncul) dalam akun *twitter* Shopee Indonesia (@*ShopeeID*) dengan judul “**Analisis Text Clustering Akun Fanpage Shopee Dengan Komentar Followers Menggunakan Tools Orange Data Mining**”.

2. METODOLOGI PENELITIAN

2.1. Metode Penelitian

Pada dasarnya *text mining* didefinisikan sebagai proses penggalian informasi dimana pengguna berinteraksi dengan kumpulan dokumen dari waktu ke waktu dengan menggunakan suatu alat analisis. *Text mining* mencari informasi dari sumber-sumber data melalui identifikasi dan eksplorasi pola tertentu, dalam kasus ini sumber data adalah kumpulan dokumen dengan pola yang ditemukan pada data teks yang tidak berstruktur. Praproses dari *text mining* sendiri berpusat pada identifikasi dan ekstraksi fitur representatif untuk dokumen *Natural Language* [4].

Proses *text mining* membutuhkan penyusunan teks masukan berdasarkan tata bahasa, yang diikuti dengan menggali pola dari data yang sudah terstruktur, evaluasi dan interpretasi hasil. Proses ini biasanya digunakan untuk pengklasifikasian, pengelompokkan, analisis makna, pengambil kesimpulan dari dokumen dan pemodelan hubungan objek yang berupa kata. Berikut merupakan tahapan dalam *text mining*:

1. Information Retrieval

Yaitu tahapan untuk memperoleh dokumen yang sesuai dengan permintaan peneliti atau yang sesuai dengan permasalahan.

2. *Natural Language Processing*

Yaitu tahapan untuk mentransformasi kata-kata yang terdapat dalam dokumen yang telah diperoleh sebelumnya. Dimana dari dokumen awalnya yang tidak terstruktur menjadi lebih terstruktur, sehingga dapat diperoleh informasi yang lebih akurat dan berguna.

3. *Information Extraction*

Yaitu tahapan dimana informasi yang sudah diperoleh sebelumnya akan diekstrak sehingga peneliti akan lebih mudah memahami permasalahan yang diteliti melalui visualisasi yang ditampilkan.

4. *Knowledge Discovery*

Pada tahapan ini pola dari suatu dokumen mulai teridentifikasi dan pengetahuan untuk mengatasi permasalahan telah didapat.

2.2. Metode Pengumpulan Data

Dalam penelitian ini bagian-bagian penelitian akan dilakukan secara komputerisasi. Mulai dari pengambilan data dengan teknik *Web Scraping* hingga analisis *text* menggunakan *tools Orange Data Mining*.

1. *Web Scraping*

Web Scraping adalah proses pengambilan sebuah dokumen semi terstruktur dari internet, umumnya berupa halaman-halaman web dalam bahasa mark up seperti HTML atau XHTML, dan menganalisis dokumen tersebut untuk diambil data tertentu dari halaman tersebut untuk digunakan bagi kepentingan lain [3].

Aplikasi *web scraping* (juga disebut *intelligent, automated, atau autonomous agents*) hanya fokus pada cara memperoleh data melalui pengambilan dan ekstraksi data dengan ukuran data yang bervariasi. Pada kasus penelitian ini mencoba mencari informasi data yang hanya berupa konten status dan komentar (sebagai fokus) dari sebuah akun *twitter*. Metode *web scraping* dalam penelitian ini menggunakan aplikasi *scraping* yaitu *Web Scrap* yang hanya bisa digunakan di *Google Chrome* untuk *windows* yang memerlukan akun pengguna *twitter* untuk dapat mengambil informasi (konten status dan komentar) dari sebuah *link* postingan kemudian diekstrak dalam bentuk *file* format **.csv*.

2. *Orange Data Mining*

Orange Data Mining adalah software open source untuk pengolahan *Data Analytics/Data Mining*. Dibandingkan dengan software *Data Mining* lainnya, *Orange* unggul dalam hal visualisasi atau yang biasa kita sebut *visual programming*. *Orange* menyediakan banyak widget yang kita letakkan pada *canvas/drawing board* kemudian kita hubungkan dengan widget-widget lainnya. Dengan media *canvas* ini maka akan memudahkan pemakai bermain dengan data dan melakukan proses *data analytics* secara intuitif [1].

Aplikasi *Orange Data Mining* hanya fokus pada cara menganalisis *text clustering* dengan hasil data yang didapatkan melalui *scraping* dari *Web Scraper*. Pada kasus penelitian ini *Orange Data Mining* menampilkan beberapa *widget* untuk mencari informasi data kata yang dominan muncul (sering muncul) dari konten status dan komentar akun *twitter* yang akan menghasilkan tampilan *word cloud* dari *widget Orange Data Mining*.

2.3. Metode Pengolahan Data

Sebelum analisis *text* data masih berupa data mentah hasil *crawling* dari *web* dan masih mengandung beberapa simbol, aksent, dan lain-lain yang akan diproses menggunakan metode pengolahan data *Preprocess Text*. Berikut ini metode pengolahan data dalam penelitian ini:

2.4. Metode Analisis Data

Metode yang digunakan untuk penerapan *data mining* menggunakan metode *clustering* (pengelompokkan). Dalam penerapan *data mining* ini menggunakan tahapan KDD (*Knowledge Discovery in Database*) dari beberapa tahapan, yaitu: [5]

1. Data Selection

Pemilihan (seleksi) data dari sekumpulan data operasional perlu dilakukan sebelum tahap penggalian informasi dalam KDD dimulai. Data hasil seleksi yang digunakan untuk proses *data mining*, disimpan dalam suatu berkas, terpisah dari basis data operasional.

2. Pre-processing/Cleaning

Sebelum proses *data mining* dapat dilaksanakan, perlu dilakukan proses *cleaning* pada data yang menjadi fokus KDD. Proses *cleaning* mencakup antara lain membuang duplikasi data, memeriksa data yang inkonsisten, dan memperbaiki kesalahan pada data.

3. Transformation

Proses transformasi pada data yang telah dipilih, sehingga data tersebut sesuai untuk proses *data mining*. Proses dalam KDD merupakan proses kreatif dan sangat tergantung pada jenis atau pola informasi yang akan dicari dalam basis data.

4. Data Mining

Data mining adalah proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu. Teknik, metode, atau algoritma dalam *data mining* sangat bervariasi. Pemilihan metode atau algoritma yang tepat sangat bergantung pada tujuan dan proses KDD secara keseluruhan.

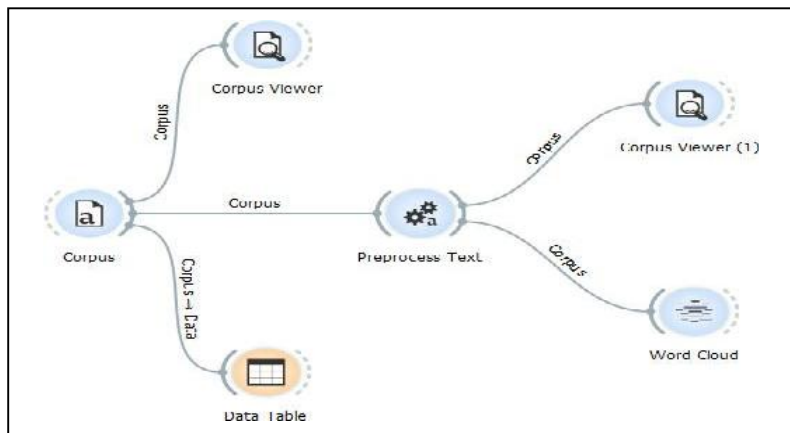
5. Interpretation/Evaluation

Pola informasi yang dihasilkan dari proses *data mining* perlu ditampilkan dalam bentuk yang mudah dimengerti oleh pihak yang berkepentingan. Tahap ini merupakan bagian dari proses KDD yang disebut *interpretation*. Tahap ini mencakup pemeriksaan apakah pola atau informasi yang ditemukan bertentangan dengan fakta atau hipotesis yang ada sebelumnya.

3. HASIL DAN PEMBAHASAN

3.1. Design Widget Text Clustering

Penggunaan *Orange Data Mining* menampilkan *Design Widget Text Clustering* yang disajikan dalam alur proses dari *Text Clustering* yang berisikan *corpus*, *corpus viewer*, *data table*, *preprocess text*, dan *word cloud*. Seperti gambar dibawah ini:



Gambar 3.1. Design Widget Text Clustering

3.2. Data Inputan (*Corpus*)

Pada penelitian ini yang menjadi data penelitian adalah *review* (status dan komentar) Shopee Indonesia di *twitter*. Status dan komentar diambil selama tiga bulan terakhir dari *postingan* akun *twitter* Shopee Indonesia. Dimana hasil akan dianalisa menjadi *cluster* dan di rekapitulasi sehingga menghasilkan kata yang dominan muncul dari status dan komentar tersebut. Dalam *Orange Data Mining* data inputan disebut juga dengan *corpus*. *Widget corpus* disebut juga dengan kumpulan dokumen yang dapat menyajikan jumlah baris kalimat, dan menentukan mana fitur yang akan dan tidak akan dimasukkan untuk keperluan analisis. *Widget corpus* dapat membaca *file-file* dari *Excell* (.xlsx), *comma-separated* (.csv) dan *native tab-delimited* (.tab).

Beberapa hal akan disajikan dalam *widget corpus* sebagai informasi laporan data yang akan dianalisis adalah: (1) Informasi tentang kumpulan data yang dimuat; (2) Fitur yang akan digunakan dalam analisis *text* (dalam hal ini hanya konten yang digunakan sebagai fitur untuk analisis *text clustering*).

3.2.1. *Corpus Viewer*

Dimaksudkan untuk melihat teks (contoh *corpus*). Itu akan selalu menampilkan *instance* dari *corpus*. Setelah proses analisis selesai *corpus viewer* digunakan untuk menampilkan kembali *corpus* (data *review*) dengan memberikan informasi berupa berapa dokumen yang ada beserta penjelasannya. Berikut ini beberapa hal yang disajikan dalam *corpus viewer* selama 3 bulan pada setiap *cluster* di status *fanpage* Shopee Indonesia:

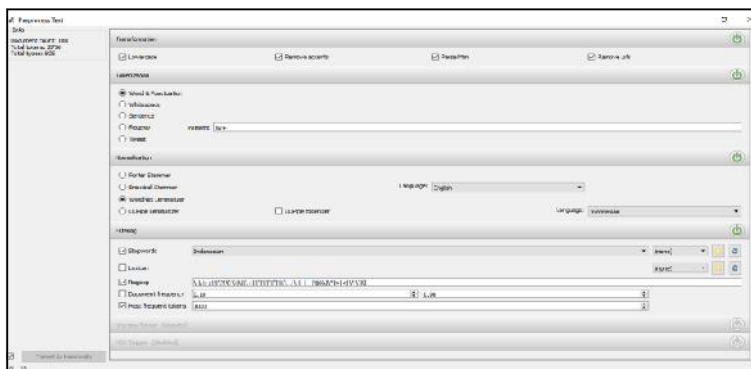
1. *Corpus Viewer Cluster 1* terdapat informasi sebanyak 108 dokumen.
2. *Corpus Viewer Cluster 2* terdapat informasi sebanyak 3 dokumen.
3. *Corpus Viewer Cluster 3* terdapat informasi sebanyak 12 dokumen.

Berikut ini beberapa hal yang disajikan dalam *corpus viewer* selama 3 bulan di komentar *followers* Shopee Indonesia:

1. *Corpus Viewer* komentar *followers* terdapat informasi sebanyak 287 dokumen.

3.3. *Preprocessing Text*

Sebelum melakukan analisis *text*, penerapan *text mining* untuk mengolah *text* dalam hal ini adalah *preprocess text*. *Text* akan dipisahkan menjadi unit yang lebih kecil (*token*), akan dilakukan *transformation*, *tokenization*, *normalization*, dan *filtering*. Yang lebih menarik langkah-langkah dalam analisis yang diterapkan secara berurutan dapat diaktifkan dan di non-aktifkan pada *Orange Data Mining* dalam *widget Preprocess Text*.



Gambar 3.2. *Preprocess Text* Pada *Orange Data Mining*

Berikut ini tahapan dalam *preprocess text* sebelum *text* dianalisis menggunakan *Orange Data Mining*:

1. *Transformation*

Langkah pertama dari *preprocess text* adalah *transformation*, yaitu mengubah data *input*. Ini berlaku untuk transformasi huruf kecil secara *default*.

- a. *Lowercase* akan mengubah semua teks menjadi huruf kecil (*Where the Vocher = where the vocher*)
- b. *Remove Accents* akan menghapus semua diakritik/aksen dalam teks. *naïve = naive*
- c. *Parse html* akan mendeteksi *tag html* dan mem-*parsing* teks saja. *<a href...> a sample of text = a sample of text*
- d. *Remove url* akan menghapus *url* dari teks. *This url http://orange.birolab.si/, = this url*

2. Tokenization

Setelah melakukan *transformation*, langkah kedua dari *preprocess text* adalah *tokenization* yaitu metode memecah teks menjadi komponen yang lebih kecil (kata, kalimat, *bigrams*).

- a. *Word & Punctuation* akan membagi teks dengan kata per kata dan membiarkan juga simbol tanda baca. *This example = (This), (example), (.)*

3. Normalization

Setelah melakukan *tokenization*, langkah ketiga dari *preprocess text* adalah *normalization* yang berlaku *stemming* dan *lemmatization text*. *Text* yang telah dipisahkan kata per kata akan menjadi sebuah *text* yang berdiri sendiri dalam sebuah kalimat. Konten status dan komentar biasanya mengandung penulisan yang kurang sempurna (*typo*). Diharapkan dengan proses normalisasi ini *text* tersebut akan diketahui maknanya dengan menggunakan:

- a. *WordNet Lemmatizer* menerapkan jaringan *sinonim kognitif* untuk *token* (kata) berdasarkan pada basis data *lexicon* (kamus) bahasa Indonesia yang besar dari NLTK (*Natural Language Toolkit*).

Data yang melalui proses *translating* secara teknis sudah diminimalkan dari istilah *typo* dimana kesalahan penulisan dari bahasa asli telah melalui proses normalisasi terlebih dahulu. Tujuan dari normalisasi pada proses ini adalah untuk melakukan *double normalization* pada data sebelum dan sesudah proses *translatating* dilakukan pada *Orange Data Mining* menggunakan perintah *WordNet Lemmatizer* agar diharapkan menghasilkan kata-kata yang natural.

4. Filtering

Proses terakhir pada *preprocess text* adalah *filtering* yaitu menghapus atau menyimpan pilihan kata. Disini adalah proses dimana proses *filtering* kata-kata dan simbol yang tidak diperlukan dalam proses selanjutnya.

- a. *Stopwords*, menghapus *stopwords* dari teks (misalnya: menghapus ‘*and*’, ‘*or*’, ‘*this*’, ...). Ini dapat dilakukan dengan memilih bahasa yang akan disaring (*filter*). Pada NLTK *server* menyediakan *stopwords* yang dapat diunduh untuk keperluan bahasa. Namun dalam penelitian ini *stopwords* secara *default* menggunakan bahasa Indonesia, dan menggunakan *Notepad* sebagai tempat penyimpanan kata-kata yang di *stopwords*. Hal ini akan menyaring hanya kata-kata kunci yang disediakan berikut ini beberapa *stopwords* yang akan dihilangkan, *stopwords* ini terkoneksi dengan data-data pada *server* NLTK:

Tabel 3.1. Daftar *Stopwords* Dalam Setiap *Cluster* Status Shopee

<i>Cluster</i>	<i>Stopwords</i>
1	#, 00, 31, 21, ^, 3, 9, @, 1, %, 10, i, a, 000, io, 5, 90
2	#, 50, %, nise1fcger, 21, b8co9k7cin, 18, nlr3vogar6

3	#, 418wyf2jne, ibup8pfnll, 100, 50, %, whyunzband, lxokv9q2us, a, ekkqh85lin, 31, bsd, ythznwvqi, rlmlv4gcv4, mlt9xxmhn, yj03bwstj1, a1fgct5nwx, 2, gwf7ye6s88, axenhlernj
---	--

Tabel 3.2. Daftar *Stopwords* Komentar *Followers*

<i>Stopwords</i>
@, #, ^, 2, 50, 1, %, 00, 30, 3, 19, d, 5, 7, 000, 21, 4, 18, e, 10, 31, 6, ~

- b. *Regex* menghapus kata-kata yang cocok dengan ekspresi reguler. *Default* diatur untuk menghapus tanda baca. Berikut ini daftar tanda baca yang akan dihapus dari *corpus*:
\.,!,:;|!~\?\\(\|/|\\|+|'|"|"'"|'\\|'...'\\-|_|-|_|-|_|\$|&|*|>|<|\\|/|\\|/|.
- c. *Most Frequent Token*, pada dasarnya *token* merupakan *text* yang dipisahkan kata demi kata. Dalam fitur ini akan menentukan beberapa jumlah *token* terbanyak yang akan muncul dan akan dianalisis dalam sebuah (*corpus*). Dalam penelitian ini menetapkan 1000 sebagai *Most Frequent Token*.

3.4. Hasil Preprocess Text

Hasil *preprocess text* telah terbentuk *text-text* terpisah dan dapat dilihat dalam bentuk *word cloud* pada *Orange Data Mining*. Berikut ini sajian hasil *preprocess text*:

1. Hasil Preprocess Text Cluster 1

Pada *cluster 1* data yang diambil pada 3 bulan dianalisa sehingga menghasilkan *preprocess text* pada *Orange Data Mining* dan dapat dilihat dalam *Word Cloud* berikut ini:



Gambar 3.3. Hasil *Word Cloud Cluster 1* Status Shopee

Berikut ini informasi *review* pengguna dengan 19 besar kata-kata dengan kemunculan terbanyak sebagai berikut:



Weight	Word
22	ku
20	mi
16	yuk
17	gak
27	kuis
20	lau
22	uon
22	ya
21	kuisshopeeive
21	jan
21	retweet
21	hai
19	live
18	lelek
18	yuhuu
16	twere
16	kak
15	yuhuuu
12	munggu

Gambar 3.4. Most Frequent Word Cluster 1 Status Shopee

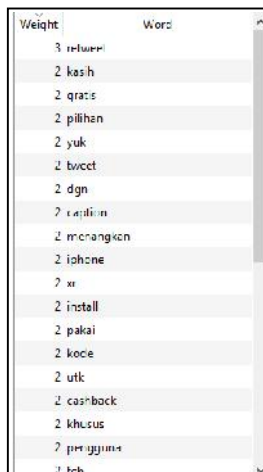
2. Hasil Preprocess Text Cluster 2

Pada cluster 2 data yang diambil pada 3 bulan dianalisa sehingga menghasilkan *preprocess text* pada *Orange Data Mining* dan dapat dilihat dalam *Word Cloud* berikut ini:



Gambar 3.5. Hasil Word Cloud Cluster 2 Status Shopee

Berikut ini informasi *review* pengguna dengan 19 besar kata-kata dengan kemunculan terbanyak sebagai berikut:



Weight	Word
3	retweet
2	kasih
2	gratis
2	pilihan
2	yuk
2	tweet
2	dgn
2	caption
2	menangkan
2	iphone
2	xi
2	install
2	pakai
2	kode
2	utk
2	cashback
2	khusus
2	pengguna
2	feb

Gambar 3.6. Most Frequent Word Cluster 2 Status Shopee

Weight	Word
29	vota
27	apik
27	apikasi
26	nih
25	yuk
24	wih
21	kasih
20	hai
22	inJamae
21	mohon
21	kuis
20	promo
24	ge
20	tau
20	teknopedia
19	terima
19	jula
19	votedkaluppee
19	the

Gambar 3.10. *Most Frequent Word* Komentar Followers Shopee

Dalam hasil tampilan diatas yaitu *word cloud* (awan kata) dengan frekuensi jumlah kata terbanyak. Semakin banyak frekuensi kemunculan suatu kata, maka semakin besar pula ukuran hurufnya pada *word cloud*. *Word cloud* merupakan variasi untuk menampilkan hasil dari tahapan *preprocess text*. Warna dari kata membuat tampilan lebih menarik, dan mudah dipahami, beberapa gambar diatas adalah hasil tahapan *preprocess text* dari data-data yang sebelumnya berisi baris kalimat status dan komentar Shopee Indonesia selama tiga bulan.

5. Hasil Analisis Text Clustering

Hasil analisis *text clustering* akun *fanpage* Shopee Indonesia terbagi menjadi 3 kelompok. Adapun contoh hasil pengelompokkan data adalah sebagai berikut:

Tabel 4.3. Hasil Pengelompokkan Data Konten Status Shopee

No.	Konten	Cluster
1.	Yuhuu, Sobat Shopee! Spesial di Kuis Shopee Day hari ini, #KuisShopeeLIVE akan ada 4x sehari Selain itu kamu bisa mendapatkan total hadiah hingga 110 JUTA! Cek di sini http://shopee.co.id/elektronik-sale Retweet kalau kamu mau kisi-kisi kuis hari ini! #ShopeeID #Shopee313Sale pic.twitter.com/JjFPy0Y85r	1
2.	#PilihShopee sebagai pilihan terbaikmu! Yuk retweet tweet ini dgn caption #PilihShopee dan menangkan iPhone XR! Lalu install Shopee dan pakai kode: PILIHSHOPEE utk Cashback 50% khusus pengguna baru Hanya hingga 18 Feb 2019, jangan ketinggalan!Cek: http://shopee.co.id/pilihshopee pic.twitter.com/NIR3VOGaR6	2
3.	Yuhuu, Sobat Shopee! Sekarang saatnya Tebak Judul Film nih. Yuk, langsung aja tebak sebanyak-banyaknya! #ShopeeID pic.twitter.com/aXeNhLErnJ	3

Dari hasil pengelompokkan data konten status Shopee diatas, maka peneliti melakukan analisis dari konten status Shopee Indonesia dengan komentar *followers*-nya untuk mendapatkan hasil kata yang dominan muncul menjadi sebuah kategori tema pembahasan yang menjadi topik yang banyak diperbincangan oleh penggunanya. Dapat dilihat dari tabel dibawah:

Tabel 4.4. Hasil Pengelompokkan Data Komentar *Followers* Shopee

No. Cluster	Akun	Nama	Komentar
1	@du2kpalingdepan	EL	Bulan Februari lalu ikut 7 lomba (Foto,nulis,kuis). 5 diantaranya kalah. 1 belum diumumkan. 1 lagi, Alhamdulillah menang. Yang penting banyak usaha. Pasti ada hasilnya. Btw, thankyou @ShopeeID pic.twitter.com/m0ROgdrTaT
2	@benzbara_	bernard batubara	ada yg lagi butuh hp baru? saya kasih tahu nih cara gampangnya. gapake duit. retweet tweet @ShopeeID yang ini + comment yg kreatif pake hestek #PilihShopee. kamu bisa dapet iPhone XR! belanja juga di Shopee pake kode PILIHSHOPEE dapet cashback 50%, sampai 18 Feb besok lho. https://twitter.com/ShopeeID/status/1096824564573556736 ...
3	@OfficialGTVID	GTV Official	Ayo dukung kontestan favoritmu untuk jadi Pemenang The Voice Indonesia. Download Shopee dan vote kontestan favoritmu di Shopee! Ada hadiah uang tunai untuk 10 voter beruntung setiap minggunya & 2 paket liburan ke Bali! @ShopeeID #VoteDiShopee #TheVoiceGTV #MakeItLouder pic.twitter.com/fFp5RGXvQ7

Dimana dari tabel diatas peneliti mengambil beberapa contoh hasil *scraping* komentar dari *followers*-nya Shopee Indonesia yang menjadi kategori tema pembahasan pada setiap *cluster*.

Dari hasil rekapitulasi *preprocess text* dan hasil analisis *text clustering* kategori topik pembicaraan pada status dan komentar dari *twitter* Shopee Indonesia (@ShopeeID), maka hasil pembahasan pada penelitian ini sebagai berikut:

1. Pada *cluster* 1 dari status *fanpage* Shopee Indonesia membahas konten tentang kuis sehingga peneliti mengambil contoh dari hasil data *scraping* komentar *followers* dimana pengguna telah memenangkan hadiah dari kuis Shopee Indonesia tersebut.
2. Pada *cluster* 2 dari status *fanpage* Shopee Indonesia membahas konten tentang *retweet* yang memberikan *handphone* sehingga menarik perhatian *followers*-nya untuk berkomentar sebanyak-banyaknya dengan *retweet* yang memberikan *handphone*.
3. Pada *cluster* 3 dari status *fanpage* Shopee Indonesia membahas konten tentang sebuah tebakan dari film sehingga hal tersebut menarik perhatian *follower*-nya untuk ikut berkomentar memberikan *vote* dari hasil tebakan film tersebut.

4. KESIMPULAN

Dari hasil analisis *text clustering* yang dilakukan pada status dan komentar *fanpage* Shopee Indonesia didapat kesimpulan sebagai berikut:

1. Dari data *fanpage* Shopee Indonesia dapat dianalisa untuk memberikan gambaran pesan yang sering muncul antara status dan komentar dari *followers* Shopee Indonesia sehingga dapat menjadi topik pembicaraan pada setiap *cluster* yang dibuat.
2. Dengan metode *clustering* yang menggunakan *k-means* menghasilkan 3 *cluster* sehingga menjadi sebagai topik pembicaraan *fanpage* Shopee Indonesia tentang kuis, *retweet* yang memberikan *handphone*, dan tebak-tebakan.
3. Dari analisis *text clustering* pada komentar dapatlah topik pembicaraan pada *fanpage* Shopee Indonesia tentang hadiah, *iphone*, *vote*.

DAFTAR PUSTAKA

- [1] Alamsyah, Andri. 2015. More than words: Social networks “Data Analytics Menggunakan Orange”. Diakses 1 April 2019, dari <https://andrya.staff.telkomuniversity.ac.id/data-analytics-menggunakan-orange/>
- [2] Indraloka, D. S., & Santosa, B. (2017). Penerapan Text Mining untuk Melakukan Clustering Data Tweet Shopee Indonesia. *Jurnal Sains dan Seni ITS*, 6(2), A51-A56.
- [3] Josi, A., & Abdillah, L. A. (2014). Penerapan teknik web scraping pada mesin pencari artikel ilmiah. *arXiv preprint arXiv:1410.5777*.
- [4] Pratama, J. A., Sunengsih, N., & Suherman, M. (2018). ANALISIS KLASSTER PADA DOKUMEN TEKS OPINI PENGGUNA TWITTER TERHADAP KASUS MIRAS OPLOSAN MENGGUNAKAN METODE K-MEANS. *Jurnal Statistika Universitas Muhammadiyah Semarang*, 6(1).
- [5] Yunita, F. (2018). PENERAPAN DATA MINING MENGGUNAKAN ALGORITMA K-MEANS CLUSTRING PADA PENERIMAAN MAHASISWA BARU. *SISTEMASI*, 7(3), 238-249.
- [6] Negara, E. S., Andryani, R., & Saksono, P. H. (2016). Twitter data analytics: geospatial data extraction and analysis, *J. INKOM*, 10(1), 27-36.
- [7] Antoni, D., Negara, E. S., & Suweno, S. (2015). Ekstraksi Data Geo-Spatial Twitter (Studi Kasus: Badan Penyelenggara Jaminan Sosial Kesehatan).
- [8] Negara, E. S. (2017). Kajian Terhadap Tools dan Framework Social Media Analytics Untuk Pemanfaatan Data Social Media Dalam Penelitian Ilmu Sosial. *Jurnal Teknologi Technoscientia*, 9(2), 132-138.
- [9] Andryani, R., Negara, E. S., & Triadi, D. (2019). Social Media Analytics: Data Utilization of Social Media for Research. *Journal of Information Systems and Informatics*, 1(2), 193-205.