e-ISSN : 2654-3168
p-ISSN : 2654-3257

Penerapan Text Mining dan Metode DBSCAN untuk Clustering Data Tweet E-Commerce

Application of Text Mining and DBSCAN Method for Clustering E-Commerce Tweet Data

Alven Safik Ritonga[1], Isnaini Muhandhis[2]
[1] Universitas Wijaya Putra, Surabaya
[2] Universitas Wijaya Putra, Surabaya
Corresponding author : alvensafik@uwp.ac.id.

Abstrak

Perkembangan teknologi dan informasi pada saat ini membuat pelaku usaha atau e-commerce beralih beriklan melalui website, sosial media; Facebook, Instagram, Twitter. Dengan beriklan di sosial media misalnya di Twitter, pelaku usaha harus jeli untuk memberikan tweet yang sharable dimana para follower akan secara sukarela membagikan konten seperti foto, video, diskon atau kuis dan pertanyaan yang dapat mendongkrak penjualan. Tujuan penelitian ini adalah membantu pelaku usaha atau e-commerce untuk mengetahui jenis konten tweet yang banyak dilakukan retweet oleh followers, sehingga konten tersebut sebagai sarana untuk melakukan promosi kepada pengguna Twitter. Untuk mendapatkan konten tersebut dilakukan clustering data tweet dari Twitter dengan menggunakan Text Mining dan metode Density-Based Spatial Clustering of Applications with Noise (DBSCAN). Metode ini membentuk cluster dari data-data yang saling berdekatan, sedangkan data yang saling berjauhan tidak akan menjadi anggota cluster. Penentuan jumlah cluster terbaik dilakukan dengan menggunakan metode Silhouette coefficient. Penelitian ini menggunakan data teks yang diambil dari akun Twitter @bliblidotcom. Hasil penelitian ini mendapatkan jumlah clustering yang terbaik berdasarkan Silhouette coefficient adalah lima cluster. Tweet yang retweet terbanyak adalah tweet opporeno di cluster 4 dan 5. Menggunakan hasil cluster tersebut Blibli Indonesia terbantu untuk melakukan promo apa yang paling disuka oleh pelanggannya.
Kata Kunci : *Text Mining, Twitter, DBSCAN, E-commerce, Clustering*

Abstract

The development of technology and information at this time makes business or e-commerce actors switch to advertising through websites, social media; Facebook, Instagram and Twitter. By advertising on social media, for example on Twitter, business actors must be observant to provide sharable tweets where followers will voluntarily share content such as photos, videos, discounts or quizzes and questions that can boost sales. The purpose of this study is to help business or e-commerce actors to find out the type of tweet content that is mostly retweeted by followers, so that the content is used as a means to promote Twitter users. To get the content, tweet data clustering from Twitter was carried out using Text Mining and the Density-Based Spatial Clustering of Applications with Noise (DBSCAN) method. This method forms clusters of data that are close to each other, while data that are far from each other will not become cluster members. Determination of the best number of clusters is done by using the Silhouette coefficient method. This study uses text data taken from the Twitter account @bliblidotcom. The results of this study get the best number of clustering based on the Silhouette coefficient is five clusters. The tweets with the most retweets were opporeno's tweets in clusters 4 and 5. Using the results of these clusters, Blibli Indonesia was helped to do what promos their customers liked the most.

Keywords : *Text Mining, Twitter, DBSCAN, E-commerce, Clustering*

PENDAHULUAN

Perkembangan dunia industri pada era industri 4.0 sekarang ini, sangat pesat. Ini terjadi karena perkembangan teknologi dan informasi. Terjadi perubahan cara

e-ISSN : 2654-3168
p-ISSN : 2654-3257

orang memproduksi barang dan jasa, contohnya kalau dulu orang belanja langsung datang ke toko, atau ke pusat perbelanjaan, tapi sekarang orang banyak beralih belanja dengan online dengan alasan lebih murah dan praktis. Contoh lain lagi, kalau dulu perusahaan-perusahaan kalau mau mempromosikan produk menggunakan billboard, baliho, brosur, televisi dan lain-lain, tapi sekarang banyak yang beralih beriklan melalui website, sosial media; Facebook, Instagram, Twitter.

Pelaku usaha atau e-commerce harus mencari tahu percakapan yang menarik dan ramai diperbincangkan oleh para nietizen, khususnya yang terkait produk atau pun kompetitor. Kemudian secara aktif dan terjadwal memberikan konten, tidak hanya terkait produk tetapi juga tips serta informasi yang sedang ngetren dan menarik bagi follower. Di samping itu, pelaku usaha juga harus berinteraksi dengan follower, aktif membalas dan memention mereka untuk berdiskusi. Pelaku usaha harus jeli untuk memberikan tweet yang sharable dimana para follower akan secara sukarela membagikan konten seperti foto, video, diskon atau kuis dan pertanyaan yang dapat mendongkrak penjualan. Semakin dikenalnya pelaku usaha di twitter, promosikan produk yang mampu membantu mendorong pertumbuhan jumlah follower dan menjangkau lebih banyak pelanggan.

Sebagian besar data tweet pada *Twitter* disimpan dalam bentuk teks, maka untuk mengetahui jenis konten dari data tweet tersebut perlu dilakukan analisis *text mining*. Salah satu teknik analisis dalam text mining adalah text clustering. Komentar (tweet) dan retweet yang dilakukan oleh followers berbentuk teks dapat dikelompokkan menjadi beberapa kluster berdasarkan kesamaan dan kemiripan konten (Putri, R.K., 2019). Dalam penelitian ini metode untuk melakukan clustering adalah metode Density-Based Spatial Clustering of Applications with Noise (DBSCAN). Algoritma pengelompokan yang didasarkan pada kepadatan (density) data. Metode ini membentuk cluster dari data-data yang saling berdekatan, sedangkan data yang saling berjauhan tidak akan menjadi anggota cluster. Penentuan jumlah klaster terbaik dilakukan dengan menggunakan metode Silhouette coefficient.

Berdasarkan uraian di atas, peneliti ingin mengkaji perilaku konsumen dalam menggunakan teknologi informasi, terutama sosial media. Dalam penelitian ini, peneliti menerapkan text mining dan Metode clustering DBSCAN untuk mengetahui kebiasaan orang memberikan komentar mengenai suatu konten e-commerce di sosial media Twitter.

1. Text Mining

Text Mining didefinisikan sebagai proses intensif pengetahuan yang dilakukan pengguna berinteraksi dengan kumpulan dokumen dari waktu ke waktu dengan menggunakan seperangkat alat analisis. Sama seperti data mining, text mining berusaha untuk mengekstrak informasi yang berguna dari sumber data melalui identifikasi dan eksplorasi pola yang menarik. Perbedaannya yaitu terdapat pada sumber data yang digunakan. Sumber data text mining berupa teks tidak terstruktur, sedangkan data mining menggunakan data terstruktur. Sehingga text mining merupakan sebuah penemuan baru dari informasi yang belum diketahui dengan mengekstrak informasi dari sumber tertulis (Feldman, R., 2007).

e-ISSN : 2654-3168
p-ISSN : 2654-3257

2. 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:

a. 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). Sebagai contoh, teks "ADVERTISING", "Advertising", "AdverTising", atau"advertising", tetap dibaca yang sama yaitu "advertising".
b. Tokenizing yaitu memotong sebuah kalimat berdasarkan tiap kata yang menyusunnya.
c. Filtering yaitu menghilangkan kata-kata yang tidak mengandung makna atau sering disebut sebagai stopword. Stopword merupakan kosa kata yang tidak memiliki ciri dari suatu dokumen.
d. 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. Teknik stemming diperlukan selain untuk memperkecil jumlah indeks yang berbeda dari suatu dokumen, juga untuk melakukan pengelompokan kata-kata lain yang memiliki kata dasar dan arti yang serupa namun memiliki bentuk yang berbeda karena mendapat imbuhan yang berbeda.
e. 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.

3. Term Frequency-Inverse Document Frequency (TF-IDF)

Term Frequency – Inverse Document Frequency (TF-IDF) adalah teknik pembobotan berbasis statistik yang sering diterapkan di berbagai permasalahan penggalian informasi. Namun, secara umum TF-IDF tidak banyak dikenal sebagai algoritma untuk peringkasan teks otomatis (Wahid, D.H. 2016). Pada peringkasan teks otomatis menggunakan TF-IDF, ide dasarnya adalah memberikan bobot pada setiap kalimat dalam sebuah dokumen. Setelah masing-masing kalimat diberikan bobot, kalimat akan diurutkan berdasarkan bobot dimana kalimat k teratas dengan bobot paling besar akan dipilih sebagai hasil akhir ringkasan. Bobot kalimat diperoleh dari penjumlahan bobot term pada sebuah kalimat, dimana term dapat berupa kata, frasa atau tipe sintatik lainnya.

4. Density-Based Spatial Clustering of Applications with Noise (DBSCAN)

DBSCAN merupakan algoritma yang masuk dalam kategori density-based clustering, yaitu proses pembentukan cluster dilakukan berdasarkan tingkat kedekatan/kerapatan jarak antar obyek dalam dataset tersebut. Beberapa kelebihan yang dimiliki algoritma DBSCAN dibandingkan dengan algoritma clustering yang lain diantaranya adalah kemampuannya dalam mendeteksi outlier/noise. Hal tersebut dikarenakan konsep density-based yang digunakan, yaitu obyek yang tidak memiliki kedekatan jarak dengan obyek yang lainnya akan dikenali sebagai outlier. Selain itu, dalam DBSCAN tidak perlu mengetahui jumlah cluster (k) yang akan terbentuk, sebagaimana yang ada pada k-Means dan k-Medoids. DBSCAN juga

e-ISSN : 2654-3168
p-ISSN : 2654-3257

mampu mengenali bentuk cluster yang tidak beraturan yang sulit dikenali oleh beberapa algoritma clustering populer lainnya. DBSCAN menentukan sendiri jumlah cluster yang akan dihasilkan sehingga kita tidak perlu lagi untuk menentukan jumlah cluster yang diinginkan, tetapi memerlukan 2 input lain, yaitu (Putra, I.M.S., 2013):

1) MinPts: minimal banyak items dalam suatu cluster.
2) Eps: nilai untuk jarak antar-items yang menjadi dasar pembentukan neighborhood dari suatu titik item.

Neighborhood yang terletak di dalam radius (ε) disebut ε-neighorhood dari objek data. Jika ε-neighborhood dari suatu objek berisi paling sedikit suatu angka yang minimum, MinPts dari suatu objek, objek tersebut disebut core object. Neighborhood dari border points berisi jauh lebih sedikit items daripada neighborhood dari core points. Suatu border point bisa jadi termasuk ke dalam lebih dari 1 core object.

METODE

Pengumpulan data yang dilakukan melalui beberapa tahapan sebagai berikut:

1. Pengumpulan Data Twitter (Crawling)

a. Pada tahapan pengumpulan data, peneliti mengambil data melalui Twitter yaitu berupa data tweet, jumlah retweet, jumlah like, dan tanggal unggahan yang ditujukan kepada akun Twitter Blibli.com Indonesia yaitu @bliblidotcom dengan bantuan Twitter API (Application Programming Interface) dan R Studio.
b. Data tweet dalam bentuk teks yang diambil adalah tweet pada akun Twitter @bliblidotcom, dari tanggal 18 Agustus 2021-26 Agustus 2021.

2. Preprocessing Data

Berikut adalah tahapan-tahapan dari Preprocessing

a. Case folding
   Tahapan 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).
b. Tokenizing
   Tahapan tokenizing, yaitu tahap pemisahan tweet berdasarkan karakter spasi pada setiap dokumen.
c. Filtering
   Tahapan filtering, yaitu tahap membuang tweet yang tidak diperlukan dari proses tokenizing. Penghapusan tweet dilakukan dengan menggunakan stopword atau kamus kata.
d. Stemming
   Proses stemming merupakan proses mengubah bentuk dari suatu tweet menjadi bentuk kata dasarnya.
e. Tagging
   Tahapan merubah berbagai tweet dalam bentuk lampau menjadi kata awalnya, tahap ini pada umumnya dilakukan untuk teks dengan bahasa Inggris atau bahasa lainnya yang memiliki bentuk lampau.

3. Tahap Analisis dan Hasil

a. Pembobotan Kata

e-ISSN : 2654-3168
p-ISSN : 2654-3257

Pembobotan kata pada tweet bertujuan untuk memberikan bobot pada fitur kata berdasarkan frekuensi kemunculan kata. Fitur kata yang telah diberi bobot dapat digunakan untuk proses clustering. Pembobotan menggunakan metode Term Frequency – Inverse Document Frequency (TF-IDF).

b. Proses Clustering
Pada penelitian ini metode yang digunakan untuk clustering adalah, metode Density-Based Spatial Clustering of Applications with Noise (DBSCAN). Tahapan proses clustering, sebagai berikut (Han, J., 2012):
1) Menentukan parameter MinPts dan Eps.
2) Pilih tweet p secara acak
3) Menghitung jumlah tweet yang ditentukan oleh parameter radius (Eps). Jika jumlahnya mencukupi (lebih dari atau sama dengan ε), data akan ditandai sebagai inti (core point).
4) Menghitung jarak titik core point dengan point yang lain menggunakan jarak Euclidean. Berikut adalah rumus jarak Euclidean yang ditujukan pada persamaan berikut ini.

$$d_{ip} = \sqrt{\sum_{j=1}^{m} (x_{ji} - y_{jp})^2} \quad (1)$$

Keterangan:

d_{ip} = jarak Euclidean dari tweet ke-i ke pusat cluster ke-k
x_{ji} = frekuensi kemunculan kata ke-j pada tweet ke-i
y_{jp} = frekuensi kemunculan kata ke-j pada titik pusat ke-p
m = banyak data

5) Buat cluster baru dengan menambahkan tweet p ke dalam cluster.
6) Melakukan identifikasi pada tweet yang ditandai sebagai core point
7) Lanjutkan proses sampai semua point telah diproses.
8) Jika ada tweet yang tidak masuk ke dalam cluster manapun akan ditandai sebagai noise.

c. Analisis Jumlah Cluster.
Penentuan jumlah cluster terbaik dilakukan berdasarkan perhitungan nilai Silhouette Coefficient, dengan menggunakan MinPts yang ditentukan dengan berbagai kombinasi eps. Nilai Silhouette Coefficient terbesar itulah jumlah cluster terbaik.

HASIL DAN PEMBAHASAN

Pada penelitian ini, yang pertama dilakukan adalah pengambilan data dalam bentuk teks yang diambil dari akun Twitter @bliblidotdom. Pengambilan data dengan cara crawling menggunakan Twitter API, data yang didapat 2164 tweet dari tanggal 18 Agustus 2021-26 Agustus 2021. Data yang sudah diambil dari akun Twitter @bliblidotcom bisa dilihat pada tabel 1 berikut ini.

e-ISSN : 2654-3168
p-ISSN : 2654-3257

Tabel 1.
Contoh Tweet dari Akun Twitter @bliblidotcom

No Tweet	Tanggal	Text Tweet	Jumlah Retweet
1	2021-08-26 23.07.44	@bliblidotcom Rujak+kecap kombinasi apik sih!	
2	2021-08-26 21.50.12	RT @bliblidotcom: Selamat: @mbauyyie92! DM y	1,0
3	2021-08-26 21.46.21	@bliblidotcom Pas dicolokke listrik e anjlok??	,0
4	2021-08-26 21.34.54	RT @bliblidotcom: Kamu udah makan siang? Kl g	59,0
5	2021-08-26 21.34.27	@bliblidotcom @mbauyyie92 Ganti aku aja bole	,0
6	2021-08-26 21.31.57	Alhamdulillah Makasihh min blibli ?????? https:/	,0

Sumber: Twitter Akun @bliblbidotcom

1. **Pengolahan Data**
   Pada tahap ini adalah tahapan teks preprocessing , data teks yang diperoleh di atas diolah melalui beberapa tahapan yaitu;

a. Case Folding
Pada proses case folding 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). Hasil proses ini bisa dilihat contoh di tabel 2 berikut ini.

Tabel 2.
Contoh Teks Tweet Proses Case Folding

No Tweet	Teks tweet yang Sudah Case Folding
1	rujak kecap kombinasi apik
2	selamat yaaa malem ngasih racun gemes bunda ayah
3	dicolokke listrik anjlok
4	makan
5	ganti
6	alhamdulillah makasihh blibli

b. Tokenizing
Prosesnya adalah memotong sebuah kalimat berdasarkan tiap kata yang membentuknya. Contoh dari data yang kita ambil bisa dilihat pada tabel 3 berikut ini.

Tabel 3.
Contoh Hasil Proses Tokenizing

No Tweet	Hasil Tokenizing
1	Rujak
	kecap
	kombinasi
	apik
	sih

c. Filtering

Pada tahap filtering ini, prosesnya adalah menghilangkan kata-kata yang tidak mengandung makna atau dengan istilah stopword. Contoh dari data teks tweet bisa dilihat pada tabel 4 berikut ini.

Tabel 4.
Contoh Hasil Proses Filtering

No Tweet	Hasil Filtering
1	rujak
	kecap
	kombinasi
	apik

2. Pembobotan Kata

Metode pembobotan yang dipakai pada penelitian ini adalah, Term Frequency – Inverse Document Frequency (TF-IDF). Prosesnya adalah merubah data tweet menjadi matriks yang berisi frekuensi kemunculan kata pada setiap data tweet. Berdasarkan hasil pembobotan ini, jumlah kata yang menyusun 2164 tweet adalah 155 kata. Contoh matriks pembobotan bisa dilihat pada tabel 5 berikut ini. Tabel tersebut berbentuk matriks yang komponennya bobot setiap kata. Tabel ini yang digunakan untuk melakukan clustering dengan metode DBSCAN.

Tabel 5.
Contoh Hasil Pembobotan

No Tweet	launching	liga	lokal	makan	menang	menyambut	nilai	nonton	opporeno	pantau
561	0.0	0.0	0.0	0.707	0.0	0.0	0.0	0.0	0.0	0.0
562	0.0	0.0	0.0	0.0	0.408	0.0	0.0	0.0	0.408	0.0
563	0.378	0.0	0.0	0.0	0.0	0.378	0.0	0.0	0.378	0.0
564	0.0	0.0	0.0	0.707	0.0	0.0	0.0	0.0	0.0	0.0
565	0.378	0.0	0.0	0.0	0.0	0.378	0.0	0.0	0.378	0.0
566	0.0	0.447	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

3. Proses Clustering dengan DBSCAN

Pada tahap klustering, hasil pembobotan pada tabel 5 digunakan untuk proses klustering dengan metode DBSCAN. Metode ini menentukan sendiri jumlah klusternya, kita hanya menentukan dua parameter, yaitu; *Eps.*(ε) dan MinPts. Bisa dilihat pada tabel 6 hasil pengujian yang didapatkan dengan mengatur *Eps.* dan MinPts., bahwa kluster terbaik pada saat Eps. (ε) sebesar 0,55 dan MinPts. sebesar 50, karena kluster dengan parameter ini yang nilai Silhoutte Coefficient yang paling besar yaitu 0,55. dengan jumlah kluster 5 dan jumlah noise adalah 726.

Tabel 6
Pengujian Metode DBSCAN

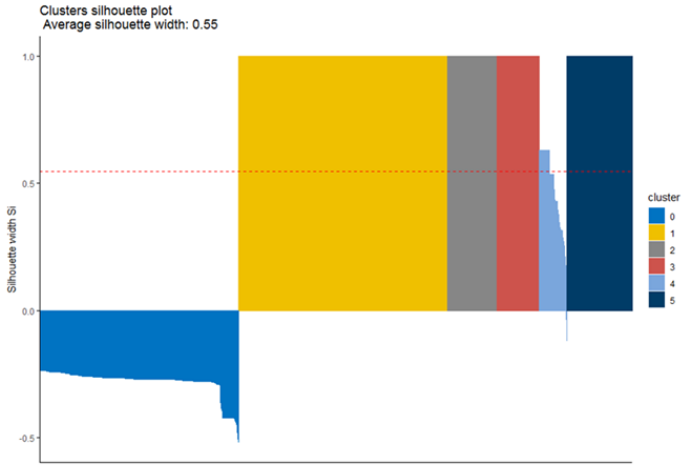
Eps. (ε)	MinPts	Silhouette Coefficient	Jumlah Cluster	Jumlah Noise
0,40	50	0,53	5	772
0,45	50	0,53	5	761

e-ISSN : 2654-3168
p-ISSN : 2654-3257

Eps. (ε)	MinPts	Silhouette Coefficient	Jumlah Cluster	Jumlah Noise
0,50	50	0,54	5	737
0,55	50	0,55	5	726
0,60	50	0,50	4	714
0,65	50	0,50	4	712
0,70	50	0,50	4	706
0,75	50	0,50	4	705
0,80	50	0,51	6	443
0,85	50	0,51	6	420
0,90	50	0,50	6	396
0,95	50	0,50	6	228

Berdasarkan nilai perhitungan Silhoutte Coefficient di atas, bisa diketahui bahwa sebanyak 4 kluster memiliki nilai positif dan 1 kluster bernilai negatif, bisa dilihat pada gambar 1 berikut ini. Grafik ini menunjukkan bahwa, kluster yang bernilai positif menandakan sebagian besar anggota kluster berada pada kluster yang tepat, demikian sebaliknya, kluster yang bernilai negatif, menandakan bahwa sebagian anggota kluster berada di kluster yang tidak tepat.

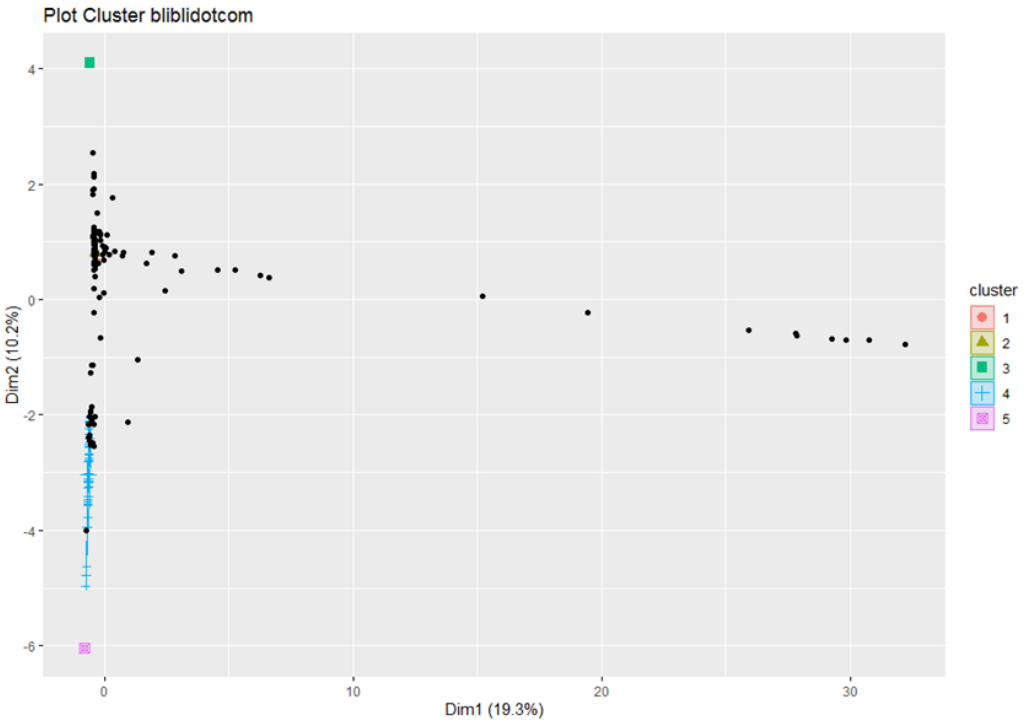
Gambar 1:
Grafik Nilai Silhouette Coeffient



Gambar penyebaran tweet bliblidotcom bisa dilihat pada gambar 2 berikut ini. Pada gambar tersebut terlihat bagaimana penyebaran tiap tweet bliblidotcom tersebar di 5 kluster, dan ada beberapa tweet yang sebagai noise, pada gambar tersebut diperlihatkan dengan titik-titik warna hitam.

e-ISSN : 2654-3168
p-ISSN : 2654-3257

Gambar 2:
Grafik Penyebaran Tweet bliblidotcom dalam Kluster



Proses klustering yang dilakukan di atas, dengan menggunakan kombinasi parameter Eps.(ε) dan Min.Pts. yang dapat dilihat di tabel 6. Maka diperoleh kluster terbaik berdasarkan nilai Silhoutte Coefficient yang terbesar yaitu kombinasi parameter Eps.(ε)=0,55 dan Min.Pts.=50. Hasil klustering tersebut dapat dilihat pada tabel 7 di bawah ini.

Tabel 7.
Hasil Klustering dengan Metode DBSCAN

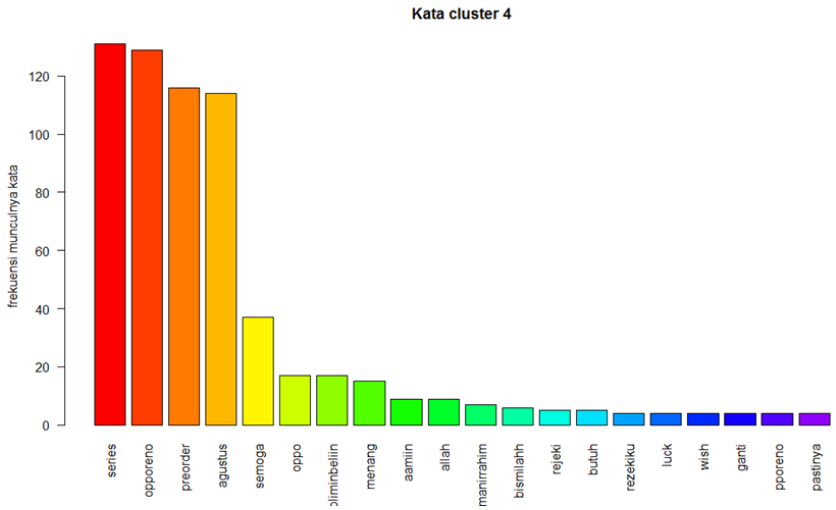
Nomor Kluster	Jumlah Tweet	Jumlah Retweet	Rata-Rata Retweet
1	762	280	0,367
2	181	188	1,038
3	155	120	0,774
4	104	570	5,480
5	238	537	2,256

Visualisasi kata-kata yang sering muncul atau yang sering di retweet follower pada tweet bliblidotcom berdasarkan masing-masing kluster yang diperoleh, ditampilkan dengan metode wordcloud dan grafik. Masing-masing kluster akan memperlihatkan wordcloud yang berbeda. Wordcloud pada gambar 3 adalah kluster 4, kata "series", "opporeno", dan "preorder" paling sering muncul di kluster 4. Setelah dilihat kata-kata selain warna hitam, dapat diambil kesimpulan awal bahwa tweet di kluster 4 adalah tweet mengenai preorder series Opporeno.

Gambar 3:
Visualisasi Wordcloud Kluster 4

Visualisasi kluster 4 juga bisa di sajikan dengan grafik bar pada gambar 4 berikut ini.

Gambar 4:
Grafik Jumlah Kemunculan Kata di Kluster 4



4. **Analisis Hasil Klustering dengan Metode DBSCAN pada Akun Twitter @bliblidotcom.**

Hasil klastering yang dihasilkan di atas, ada 2 klaster memiliki jenis konten tweet yang sama yaitu konten mengenai produk "opporeno" pada klaster 4 dan klaster 5. Penyebab kenapa konten "opporeno" tersebut berada di dua klaster yang bebeda dipengaruhi oleh kata-kata lain sebagai pembeda yaitu kata "preorder" dan kata "launching". Dari tabel 4.7 bisa dilihat bahwa meskipun kedua kluster ini memiliki konten yang hampir sama, tapi kedua kluster ini memiliki jumlah retweet dan rata-rata retweet yang tinggi. Jadi

berdasarkan hasil tersebut dapat dilihat bahwa konsumen e-commerce Blibli Indonesia lebih tertarik pada konten produk gawai yang baru yaitu opporeno baik itu preorder maupun launching produknya.

KESIMPULAN

Penerapan metode DBSCAN pada proses Text Mining untuk melakukan clustering pada data tweet e-commerce Blibli menghasilkan 5 cluster. Pada penentuan jenis konten yang paling banyak di retweet dan rata-rata retweet yang tinggi pada tiap cluster, diperoleh bahwa cluster 4 dan cluster 5, dimana konten yang paling sering retweet di 2 cluster itu adalah preorder opporeno dan launching opporeno. E-commerce Blibli Indonesia dapat menggunakan hasil Clustering ini untuk mengetahui hal apa yan paling disukai oleh followernya di twitter, dan merupakan salah satu cara yang mudah dan murah untuk melakukan advertising kepada konsumennya.

DAFTAR PUSTAKA

Feldman, R., & Sanger, J. 2007. *The text mining handbook*. New York: Cambridge University Press.

Furqon, M.T., & Muflikhah, L. 2016. "Clustering the potential risk of tsunami using density-based spatial clustering of application with noise (DBSCAN)" dalam Journal of Environmental Engineering & Suistainable Technology Vol. 3 No.1 (Hal.1-8).

Han, J., Kamber,M., & Pei, J. 2012. *Data mining concepts and techniques*. Waltham: Morgan Kaufmann Publishers.

Indrayuni, E. 2017. "Text mining dalam analisis sentimen review restoran menggunakan algoritma k-nearest-neighbor (KNN)" dalam Jurnal Teknik Informatika Stmik Antar Bangsa Vol.3 No. 2 (Hal.109-114).

Putra, I. M. S. 2013. "Segmentasi citra remote sensing laut dengan metode clustering DBSCAN" dalam Majalah Ilmiah Teknologi Elektro Vol. 12 No. 2 (Hal.16-23).

Putri, R.K., Warsito, B., & Mustafid. 2019. "Implementasi algoritma modified gustafson-kessel untuk clustering tweets pada akun twitter lazada Indonesia" dalam Jurnal Gaussian Vol.8 No.3 (Hal.285-295).

Renganathan, V. 2017. "Text mining in biomedical domain with emphasis on document clustering" dalam Healthcare Informatics Research Vol.3 No. 1 (Hal.141-146).

Sahayak, V., Shete, V., & Pathan, A. 2015. "Sentiment analysis on twitter data" dalam International Journal of Innovative Research in Advanced Engineering (IJIRAE) Vol. 2 No.1 (Hal.178-183).

Samsir, Ambiyar, Verawardina, U., Edi, F., dan Watrianthos, R., 2021, "Analisis Sentimen Pembelajaran Daring Pada Twitter di Masa Pandemi COVID-19 Menggunakan Metode Naïve Bayes" dalam Jurnal Media Informatika Budidarma Vol.5, No. 1 (Hal. 157-163).

Wahid, D.H., dan Azhari, 2016, "Peringkasan Sentimen Esktraktif di Twitter Menggunakan Hybrid TF-IDF dan Cosine Similarity" dalam IJCCS Vol. 10 No. 2 (Hal.207-218).