

Peningkatan Kualitas Layanan Perbankan Digital Melalui Pengelompokan Tweet Menggunakan DBSCAN

Syeni Agustin Ningtias¹, Alfisyahrina Hapsery^{2*},

¹Statistika, Universitas PGRI Adi Buana Surabaya, Indonesia

²Statistika, Universitas PGRI Adi Buana Surabaya, Indonesia

*Corresponding author: alfisyahrina@unipasby.ac.id

Received: 8 March 2022

Revised: 27 Agustus 2023

Accepted: 23 March 2023

ABSTRAK – Digitalisasi di bidang perbankan memungkinkan nasabah memperoleh layanan perbankan secara mandiri tanpa harus datang langsung ke bank. Layanan perbankan digital memungkinkan nasabah memperoleh informasi, berkomunikasi, mendaftar, membuka rekening, melakukan transaksi perbankan dan menutup rekening, termasuk memperoleh informasi dan transaksi lainnya di luar produk perbankan. Perbankan gencar memberikan pelayanan atau promosi melalui media sosial, salah satunya dengan menggunakan media sosial Twitter. Metode Density-Based Spatial Clustering of Applications with Noise (DBSCAN). DBSCAN dilakukan dengan menggabungkan nilai Eps dan MinPts untuk menghasilkan koefisien *silhouette* tertinggi. Nilai koefisien siluet tertinggi dari tweet layanan perbankan BRI, Mandiri dan BCA masing-masing menghasilkan 33, 14, dan 39 cluster dengan nilai Eps dan MinPts yang berbeda. Berdasarkan hasil *wordcloud* menunjukkan bahwa layanan perbankan perlu ditingkatkan dalam hal pengecekan DM pada rekening, nasabah meminta kepada admin untuk segera menanggapi keluhan terkait kartu ATM, gangguan pada mobile banking dan ada yang mengucapkan terima kasih atas layanannya yang telah disediakan.

Kata kunci– DBSCAN, Perbankan, Text Clustering, Twitter, Silhouette

ABSTRACT – Digitalization in the banking sector allows customers to obtain banking services independently without having to come directly to the bank. Digital banking services enable customers to obtain information, communicate, register, open accounts, banking transactions and close accounts, including obtaining other information and transactions outside of banking products. Banking is intensively providing services or promotions through social media, one of which is by using Twitter social media. Density-Based Spatial Clustering of Applications with Noise (DBSCAN) method. DBSCAN clustering is done by combining the Eps and MinPts values to produce the highest silhouette coefficient. The highest silhouette coefficient values from BRI, Mandiri and BCA banking service tweets produce 33, 14, and 39 clusters, respectively, with different Eps and MinPts values. Based on the results of wordcloud, it shows that banking services need to be improved in terms of checking DM on accounts, customers ask the admin to immediately respond to complaints related to ATM cards, disruptions to mobile banking and some say thank you for the services that have been provided.

Keywords– DBSCAN, Banking, Text Clustering, Twitter, Silhouette

I. PENDAHULUAN

Saat ini beberapa bidang kehidupan sedang mengalami proses migrasi teknologi digital dengan tujuan untuk mendapatkan efisiensi dan optimalisasi. Sehingga setiap industri harus siap bergerak menghadapi perubahan-perubahan yang dinamis, tak terkecuali pada industri keuangan dan perbankan. Perbankan telah menghadirkan terobosan baru untuk memanfaatkan perkembangan teknologi tersebut dengan cara melakukan layanan perbankan digital. Dimana perbankan digital memungkinkan nasabah untuk memperoleh layanan perbankan secara mandiri tanpa harus datang langsung ke bank. Bank merupakan lembaga yang bergerak dalam bidang keuangan yang memiliki wewenang untuk menghimpun dana dari masyarakat yang berbentuk simpanan. Nasabah merupakan orang yang menggunakan jasa pelayanan dari perbankan sehingga kepuasan nasabah menjadi hal yang paling penting untuk diperhatikan. Kesalahan proses transaksi, akan memicu nasabah untuk mengajukan pengaduan. Pengaduan yang dilakukan nasabah di era digital dilakukan dengan menggunakan media sosial.

Salah satu media sosial yang banyak dimanfaatkan adalah twitter. Twitter bukan hanya digunakan sebagai ajang promosi, akan tetapi bisa digunakan sebagai proses interaksi antara bank dengan nasabah. Twitter juga merupakan sarana bagi masyarakat untuk mempublikasikan berita. Berita yang dimuat dalam media sosial juga dapat dilakukan klasifikasi [3]. Perusahaan perbankan juga dapat menerima feedback berupa pendapat secara terbuka yang diberikan oleh nasabah melalui media sosial tersebut. Perbankan di Indonesia memiliki akun customer care yang disediakan sebagai layanan online untuk menanggapi tanggapan nasabah melalui media sosial twitter [9]. Bank BRI, Mandiri, BCA yang memiliki akun twitter customer care yaitu @KontakBRI, @mandiricare dan @HaloBCA, dimana ketiga bank tersebut adalah perbankan di Indonesia yang memiliki urutan tertinggi terkait dengan kenaikan modal yang ditunjukkan dari riset “the banker” [5] dan bank dengan aset yang terus meningkat di tengah pandemi covid-19 [10].

Salah satu metode statistika yang dapat diterapkan dalam analisis text mining seperti data tweet, yaitu dengan menggunakan analisis text clustering yang dapat mengelompokkan data unstructured [6]. Text clustering merupakan

pengelompokan data dengan cara memaksimalkan kesamaan antar anggota satu kelas dan meminimumkan kesamaan antar cluster [11]. Pada penelitian text clustering ini dilakukan dengan menggunakan metode Density-Based Spatial Clustering of Applications with Noise (DBSCAN). DBSCAN adalah algoritma pengelompokan yang didasarkan pada kepadatan (density) data [13]. Penelitian terkait telah dilakukan oleh Yusril dengan “Implementasi Text Mining untuk Advertising dengan Menggunakan Metode K-Means Clustering pada Data Tweets Gojek Indonesia” yang menghasilkan kesimpulan bahwa 2 cluster telah terbentuk optimal dengan konten program kuis dan pengenalan produk milik gojek Indonesia merupakan jumlah retweet terbanyak pada tiap cluster [14]. Penelitian tersebut tergolong ke dalam penelitian text mining yang membutuhkan pembobot dalam proses analisisnya yaitu TF-IDF [1], [2] lainnya dengan metode K-Means menggunakan optimasi dengan membandingkan antara Gap Statistik, silhouette coefficient, dan elbow menghasilkan metode terbaik dalam mengelompokkan adalah silhouette coefficient [12]. Penelitian lain pernah dilakukan oleh Arsih dengan membandingkan algoritma K-Means dengan DBSCAN pada data nasabah yang memiliki fasilitas kredit dengan judul “Metode Pengclusteran Berbasis Densitas Menggunakan Algoritma DBSCAN” yang menghasilkan kesimpulan bahwa kedua algoritma tersebut membentuk 2 cluster dengan algoritma DBSCAN lebih baik daripada K-Means [4][15]. Berdasarkan beberapa penelitian terdahulu, diperoleh informasi untuk metode pengelompokan terbaik yaitu dengan DBSCAN dan pemilihan optimasi dengan silhouette coefficient, sehingga keduanya akan digunakan dalam penelitian ini. Tujuan dari penelitian ini yaitu untuk mengetahui bagaimana kualitas layanan yang dilakukan perbankan kepada customer yang dilaporkan melalui akun twitter. Hasil dari klasifikasi dapat digunakan pihak terkait untuk dapat menentukan kebijakan terhadap peningkatan kualitas layanan dimasa yang akan datang. Penerapan ini dilakukan pada akun perbankan BRI, Mandiri dan BCA menggunakan Density-Based Spatial Clustering of Applications with Noise (DBSCAN).

II. METODE PENELITIAN

A. Sumber Data dan Variabel Penelitian

Data yang digunakan dalam penelitian ini adalah data yang diambil dari kumpulan tweet dari pengguna twitter di Indonesia. Akun twitter yang digunakan dalam analisis adalah customer care dari layanan BRI @KontakBRI, customer care dari layanan Mandiri @mandiricare dan customer care dari layanan BCA @HaloBCA. Data tweet diambil dari tanggal 4 Februari 2021 sampai 5 Maret 2021 dengan menggunakan twitter API. Variabel penelitian yang digunakan adalah frekuensi kemunculan kata dasar dari setiap tweet yang ditujukan kepada @KontakBRI, @mandiricare, dan @HaloBCA yang telah dilakukan preprocessing, dinotasikan sebagai variabel x dengan skala rasio.

B. Struktur Data

Struktur data yang digunakan pada penelitian ini (lihat Tabel 1) digunakan untuk layanan perbankan @KontakBRI, layanan perbankan @mandiricare dan layanan perbankan @HaloBCA.

Tabel 1 Struktur Data Penelitian

No	Tweet (i)	Kata (j)					
		x_1	x_2	x_3	x_4	x_m
1	b_1	$W_{1.1}$	$W_{2.1}$	$W_{3.1}$	$W_{4.1}$	$W_{m.1}$
2	b_2	$W_{1.2}$	$W_{2.2}$	$W_{3.2}$	$W_{4.2}$	$W_{m.2}$
3	b_3	$W_{1.3}$	$W_{2.3}$	$W_{3.3}$	$W_{4.3}$	$W_{m.3}$
:	:	:	:	:	:	:	:
i	b_i	$W_{1.i}$	$W_{2.i}$	$W_{3.i}$	$W_{4.i}$	$W_{m.i}$
:	:	:	:	:	:	:	:
n	b_n	$W_{1.n}$	$W_{2.n}$	$W_{3.n}$	$W_{4.n}$	$W_{m.n}$

Struktur data yang akan dilakukan proses clustering DBSCAN setelah dilakukan preprocessing data dimana b merupakan tweet, x merupakan variabel yang menggambarkan kata dari tweet dan W merupakan kata yang telah terboboti dengan Term Frequency-Inverse Document Frequency (TF-IDF) TF.

C. Langkah Penelitian

Langkah penelitian yang digunakan pada penelitian ini adalah sebagai berikut.

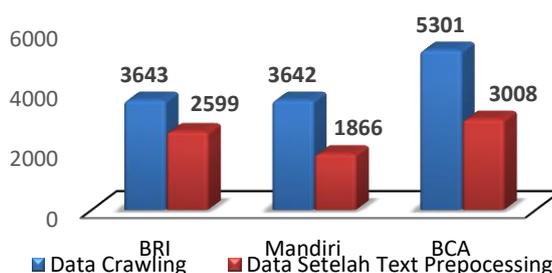
1. Mengambil data tweet dengan menggunakan Twitter API
 - a. Memasukkan keyword yang berhubungan dengan akun Twitter @KontakBRI, @mandiricare dan @HaloBCA.
 - b. Menyimpan hasil crawling ke database.
2. Melakukan text preprocessing pada data tweet layanan perbankan BRI, Bank Mandiri, dan BCA.

3. Mengubah data tweet ke dalam bentuk frekuensi kemunculan kata dengan menggunakan TF-IDF.
4. Melakukan clustering data.
 - a. Melakukan clustering dengan menggunakan algoritma DBSCAN.
 - b. Memilih hasil clustering terbaik atau optimum dengan melihat nilai silhouette coefficient terbesar untuk metode DBSCAN.
5. Melakukan visualisasi tweet dengan Word Cloud.
6. Melakukan interpretasi dan menarik kesimpulan.

III. HASIL DAN PEMBAHASAN

A. Karakteristik Data

Dari hasil crawling data, didapatkan 3.643 tweet yang ditujukan kepada akun customer care BRI, 3.642 tweet yang ditujukan kepada akun customer care Mandiri dan 5.301 tweet yang ditujukan kepada akun customer care BCA. Sebelum melakukan analisis clustering, perlu dilakukan text preprocessing agar data mentah menjadi terstruktur dan mudah untuk dilakukan analisis.



Gambar 1. Jumlah Tweet Bank BRI, Mandiri dan BCA

Berdasarkan Gambar 1 tweet yang paling banyak diberikan oleh masyarakat adalah tweet pada akun. layanan perbankan BCA. Hal ini dapat dilihat bahwa layanan customer care BCA lebih sering digunakan oleh masyarakat untuk melakukan interaksi, baik interaksi dalam hal keluhan, kritik, saran dan lain-lain. Pada proses preprocessing data, tweet yang terambil sebagai retweet, tweet yang kosong atau kalimat dan kata yang tidak memiliki makna tidak akan dimasukkan ke dalam struktur data clustering, sehingga jumlah data tweet yang digunakan sebanyak 2.599 untuk perbankan BRI, 1.866 untuk perbankan Mandiri dan 3.008 untuk perbankan BCA.

B. Clustering DBSCAN untuk Bank BRI

Clustering DBSCAN dimulai dengan melakukan perhitungan TF-IDF Berikut ini adalah hasil frekuensi kemunculan kata yang diperoleh setelah dilakukannya text preprocessing pada akun customer care BRI.

Tabel 2. Frekuensi Layanan Perbankan BRI

Tweet ke-	Kata						
	admin	..	cek	..	otp	..	whatsapp
1	1	..	0	..	0	..	0
2	1	..	1	..	0	..	0
:	:		:		:		:
1485	1	..	0	..	0	..	0
:	:		:		:		:
2599	1	..	1	..	0	..	0

Tabel 2 menunjukkan frekuensi kemunculan setiap kata pada tiap tweet yang didapatkan dari proses text preprocessing, dengan jumlah kata sebanyak 117 dan kata tersebut merupakan variabel penelitian yang akan digunakan. Berikut adalah hasil perhitungan TF-IDF.

Tabel 3. Perhitungan TF-IDF Layanan Perbankan BRI

Tweet ke-	Kata						
	admin	..	cek	..	otp	..	whatsapp
1	0,2204	..	0	..	0	..	0
2	0,2198	..	0,2625	..	0	..	0
:	:		:		:		:
1485	0,3563	..	0	..	0	..	0
:	:		:		:		:
2599	0,5310	..	0	..	0	..	0

Tabel 3 menunjukkan bahwa kata “cek” pada tweet kedua memiliki bobot sebesar 0,2625. Hal ini menunjukkan bahwa kata “cek” pada tweet tersebut memiliki frekuensi kemunculan yaitu $1 \times 0,2625 = 0,2625$. Begitupula dengan kata-kata dasar yang lainnya memiliki cara perhitungan TF-IDF yang sama pula. Berikut merupakan visualisasi word cloud yang dibentuk berdasarkan bobot TF-IDF.



Gambar 2. Visualisasi Word Cloud Perbankan BRI

Gambar 2 menunjukkan bahwa nasabah BRI meminta admin customer care untuk melakukan pengecekan dan tindak lanjut terkait dengan dm yang diberikan, hal tersebut terlihat dari kata “admin”, “cek”, “dm” dan “tolong”. Nasabah juga memberikan ungkapan terimakasih karena telah menyelesaikan masalahnya dan memberikan informasi. Kepuasan dan kendala nasabah terkait dengan aplikasi brimo, transaksi rekening, topup, transfer yang gagal dan lain-lain juga disampaikan oleh nasabah melalui twitter customer care BRI. Menunjukkan kata yang paling sering muncul pada akun layanan perbankan BRI adalah kata “admin” dan “dm” sebanyak 1.154 dan 1.039 kali kemunculan, jumlah kemunculan kata tersebut menunjukkan bahwa banyaknya nasabah yang meminta agar customer service melakukan pengecekan pada fitur Direct Message.

Dua parameter hyperparameter yaitu nilai parameter Eps dan MinPts, dimana cluster akan terbentuk ketika memiliki sedikitnya MinPts data pada radius Eps. Pada analisis ini akan dilakukan looping untuk mencari kombinasi Eps dan MinPts yang menghasilkan silhouette coefficient tinggi. Kombinasi yang digunakan adalah nilai Eps sebesar 0.1, 0.2, hingga 1 dan nilai MinPts sebesar dari 10 hingga 100. Berikut merupakan tabel kombinasi nilai parameter Eps dan MinPts yang menghasilkan nilai silhouette coefficient tertinggi. Nilai silhouette merupakan metode untuk mengetahui jumlah kelompok optimum dari hasil cluster. Terdapat beberapa metode diantaranya gap statistic, silhouette dan elbow, diantara ketiganya yang menunjukkan hasil terbaik adalah silhouette [8].

Tabel 4. Kombinasi Nilai Parameter Eps dan MinPts pada Layanan Perbankan BRI

MinPts	Eps	Silhouette Coefficient	Noise	Jumlah Cluster
10	0,8	0,01947	417	15
11	0,8	0,03339	457	14
12	0,8	0,03627	504	13
13	0,8	0,03641	554	13
14	0,8	0,03769	614	12
:	:	:	:	:
15	0,5	0,23697	1494	25
12	0,5	0,23698	1410	31
17	0,5	0,23937	1528	23
11	0,5	0,23994	1394	32
10	0,5	0,24065	1375	33

Tabel 4 menunjukkan bahwa kombinasi nilai parameter Eps dan MinPts yang menghasilkan nilai silhouette coefficient tinggi terdapat pada MinPts sebesar 10 dan Eps sebesar 0,5. Oleh karena itu, jumlah cluster optimum yang diperoleh sebanyak 33 cluster dengan 1.224 tweet yang dapat dikelompokkan dan 1.375 tweet yang tidak dapat dikelompokkan atau sebagai noise.

Tabel 5. Jumlah Anggota Cluster pada Layanan Perbankan BRI

Cluster	Jumlah Anggota	Cluster	Jumlah Anggota	Cluster	Jumlah Anggota
1	20	12	121	23	35
2	32	13	20	24	33
3	275	14	59	25	41
4	26	15	29	26	18
5	15	16	16	27	36
6	46	17	24	28	28
7	14	18	13	29	10
8	15	19	17	30	31
9	22	20	61	31	15
10	24	21	18	32	11
11	26	22	61	33	12

Tabel 5 menunjukkan bahwa cluster yang terbentuk pada data layanan perbankan BRI memiliki jumlah anggota sebanyak 20 untuk cluster pertama, 32 untuk cluster kedua dengan jumlah anggota terbanyak ada pada cluster 3 yaitu sebanyak 275 tweet.

Tabel 6. Anggota Cluster Pertama pada Layanan Perbankan BRI

Tweet ke-	Tanggapan Nasabah	Tweet ke-	Tanggapan Nasabah
2	terimakasih kak	1586	terimakasih kak
10	terimakasih kak	1625	terimakasih kak
21	terimakasih kak	1936	terimakasih kak
25	terimakasih kak	1950	terimakasih kak
218	terimakasih kak	1955	terimakasih kak
654	terimakasih kak	1962	terimakasih kak
897	terimakasih kak	1987	kak terimakasih
934	kak terimakasih	2059	terimakasih kak
1334	kak terimakasih	2094	terimakasih kak
1511	kak terimakasih	2113	terimakasih kak

Tabel 6 menunjukkan bahwa kata yang muncul pada cluster pertama adalah “terimakasih” dan “kak”, dimana kata tersebut merupakan kata yang muncul dari 20 tweet di cluster pertama. Beberapa cluster yang terbentuk mempunyai makna yang sama, dimana cluster 3, 4, 6, 8, 9, 10, 12, 15, 16, 19, 22 dan cluster 33 menunjukkan bahwa nasabah ingin meminta customer service atau admin untuk memeriksa, membaca, merespon nasabah melalui fitur dm. Cluster 1, 2, 13, 14, 17 dan cluster 30 menunjukkan ungkapan terimakasih nasabah karena admin telah memberikan informasi, tips, edukasi atau mengatasi masalah nasabah. Sedangkan cluster 5, 7, 18, 20, 21, 29 dan cluster 31 menunjukkan keluhan nasabah terhadap aplikasi mobile banking dan brimo yang error.

C. Clustering DBSCAN untuk Bank Mandiri

Clustering DBSCAN dimulai dengan melakukan perhitungan TF-IDF Berikut ini adalah hasil frekuensi kemunculan kata yang diperoleh setelah dilakukannya text preprocessing pada akun customer care Mandiri.

Tabel 7. Frekuensi Layanan Perbankan Mandiri

Tweet ke-	Kata						
	ada	..	kartu	..	token	..	whatsapp
1	0	..	0	..	0	..	0
2	0	..	0	..	0	..	0
:	:		:		:		:
853	0	..	1	..	0	..	0
:	:		:		:		:
1866	0	..	0	..	0	..	0

Tabel 7 menunjukkan frekuensi kemunculan setiap kata pada tiap tweet yang didapatkan dari proses text preprocessing, dengan jumlah kata sebanyak 104 dan kata tersebut merupakan variabel penelitian yang akan digunakan. Berikut adalah hasil perhitungan TF-IDF.

Tabel 8. Perhitungan TF-IDF Layanan Perbankan Mandiri

Tweet ke-	Kata						
	ada	..	kartu	..	token	..	whatsapp
1	0	..	0	..	0	..	0
2	0	..	0	..	0	..	0
:	:		:		:		:
853	0	..	0,7111	..	0	..	0
:	:		:		:		:
1866	0	..	0	..	0	..	0

Tabel 8 menunjukkan bahwa kata “kartu” pada tweet kedua memiliki bobot sebesar 0,7111. Hal ini menunjukkan bahwa kata “kartu” pada tweet tersebut memiliki frekuensi kemunculan yaitu $1 \times 0,7111 = 0,7111$. Begitupula dengan kata-kata dasar yang lainnya memiliki cara perhitungan TF-IDF yang sama pula. Berikut merupakan visualisasi word cloud yang dibentuk berdasarkan bobot TF-IDF.



Gambar 3. Visualisasi Word Cloud Perbankan Mandiri

Gambar 3 menunjukkan bahwa nasabah Bank Mandiri meminta admin customer care untuk melakukan pengecekan dan tindak lanjut terkait dengan dm yang diberikan, hal tersebut terlihat dari kata “cek”, “dm”, dan “admin” yang paling menonjol dibandingkan kata-kata yang lain. Nasabah juga memberikan ungkapan terimakasih karena telah menyelesaikan masalahnya. Ketidakpuasan nasabah terhadap mobile banking yang error, proses transaksi yang gagal, dan kendala lainnya juga diungkapkan nasabah melalui twitter customer care Mandiri. Selain itu, nasabah juga menanyakan terkait dengan cara membuka rekening baru, cara mengurus buku tabungan yang hilang dan lain-lain.

Berikut merupakan hasil analisis clustering menggunakan metode DBSCAN pada layanan perbankan Mandiri. Terdapat 2 parameter hyperparameter yaitu nilai parameter Eps dan MinPts, dimana cluster akan terbentuk ketika memiliki sedikitnya MinPts data pada radius Eps. Pada analisis ini akan dilakukan looping untuk mencari kombinasi Eps dan MinPts yang menghasilkan silhouette coefficient tinggi. Kombinasi yang digunakan adalah nilai Eps sebesar 0.1, 0.2, hingga 1 dan nilai MinPts sebesar dari 10 hingga 100. Berikut merupakan tabel kombinasi nilai parameter Eps dan MinPts yang menghasilkan nilai silhouette coefficient tertinggi.

Tabel 9. Kombinasi Nilai Parameter Eps dan MinPts pada Layanan Perbankan Mandiri

MinPts	Eps	Silhouette Coefficient	Noise	Jumlah Cluster
17	0,8	0,01990	425	7
13	0,8	0,02020	357	4
14	0,8	0,02188	369	5
15	0,8	0,02260	384	5
12	0,8	0,02600	348	3
:	:	:	:	:
11	0,4	0,16141	1312	18
10	0,4	0,16441	1292	20
16	0,4	0,17091	1385	13
17	0,4	0,17091	1385	13
15	0,4	0,17139	1370	14

Tabel 9 menunjukkan bahwa kombinasi nilai parameter Eps dan MinPts yang menghasilkan nilai silhouette coefficient tinggi pada layanan perbankan Mandiri terdapat pada MinPts sebesar 15 dan Eps sebesar 0,4. Oleh karena itu, jumlah cluster optimum yang diperoleh sebanyak 14 cluster dengan 496 tweet yang dapat dikelompokkan dan 1370 tweet yang tidak dapat dikelompokkan atau sebagai noise.

Tabel 10. Jumlah Anggota Cluster pada Layanan Perbankan Mandiri

Cluster	Jumlah Anggota	Cluster	Jumlah Anggota
1	26	8	19
2	45	9	27
3	21	10	31
4	139	11	15
5	19	12	28
6	23	13	17
7	59	14	27

Tabel 10 menunjukkan bahwa cluster yang terbentuk pada data layanan perbankan Mandiri memiliki jumlah anggota sebanyak 26 untuk cluster pertama dan 45 untuk cluster kedua dengan jumlah anggota cluster terbanyak terdapat pada cluster 4 yaitu sebanyak 139 tweet.

Tabel 11. Anggota Cluster Pertama pada Layanan Perbankan Mandiri

Tweet ke-	Tanggapan Nasabah
6	admin cek dm terimakasih
136	cek dm admin terimakasih
165	admin cek dm terimakasih
278	cek dm admin terimakasih
281	admin cek dm terimakasih
301	admin cek dm terimakasih
364	admin cek dm terimakasih
396	cek dm admin terimakasih
:	:
1376	cek dm admin terimakasih
1396	admin cek dm terimakasih
1444	admin cek dm terimakasih
1625	cek dm admin terimakasih

Tabel 11 menunjukkan bahwa kata yang muncul pada cluster pertama adalah “admin”, “cek”, “dm”, dan “terimakasih”, dimana kata tersebut merupakan kata yang muncul dari 26 tweet di cluster pertama. Cluster 1, cluster 4 sampai cluster 8, cluster 10 sampai cluster 12, dan cluster 14 menunjukkan bahwa nasabah ingin meminta customer service atau admin untuk memeriksa, membaca, merespon dan membantu permasalahan nasabah yang ditulis melalui fitur dm atau direct Message. Cluster 2, cluster 3, cluster 9 dan cluster 13 menunjukkan bahwa nasabah mengucapkan terimakasih kepada customer service untuk informasi, edukasi dan tips yang telah diberikan dan terimakasih karena telah menyelesaikan masalah nasabah.

D. Clustering DBSCAN untuk Bank BCA

Clustering DBSCAN dimulai dengan melakukan perhitungan TF-IDF Berikut ini adalah hasil frekuensi kemunculan kata yang diperoleh setelah dilakukannya text preprocessing pada akun customer care BCA.

Tabel 12. Frekuensi Layanan Perbankan BCA

Tweet ke-	Kata						
	ada	..	bca	..	teller	..	xpresi
1	0	..	1	..	0	..	0
2	0	..	0	..	0	..	0
:	:		:		:		:
1812	0	..	1	..	0	..	0
:	:		:		:		:
3008	0	..	0	..	0	..	0

Tabel 12 menunjukkan frekuensi kemunculan setiap kata pada tiap tweet yang didapatkan dari proses text preprocessing, dengan jumlah kata sebanyak 115 dan kata tersebut merupakan variabel penelitian yang akan digunakan. Berikut adalah hasil perhitungan TF-IDF.

Tabel 15. Jumlah Anggota Cluster pada Layanan Perbankan BCA

Cluster	Jumlah Anggota	Cluster	Jumlah Anggota	Cluster	Jumlah Anggota
1	84	14	65	27	47
2	15	15	268	28	19
3	76	16	36	29	31
4	13	17	19	30	16
5	63	18	39	31	11
6	191	19	14	32	15
7	12	20	15	33	22
8	27	21	11	34	26
9	17	22	14	35	22
10	36	23	14	36	11
11	48	24	42	37	12
12	21	25	14	38	12
13	125	26	35	39	10

Tabel 15 menunjukkan bahwa cluster yang terbentuk pada data layanan perbankan BCA memiliki jumlah anggota sebanyak 84 untuk cluster pertama, 15 untuk cluster kedua dengan jumlah anggota terbanyak ada pada cluster 15 sebanyak 268 tweet.

Tabel 16. Anggota Cluster Pertama pada Layanan Perbankan BCA

Tweet ke-	Tanggapan Nasabah
5	informasi terimakasih
24	admin terimakasih informasi
91	terimakasih informasi
101	terimakasih informasi
105	terimakasih informasi
:	:
2865	terimakasih informasi admin
2869	terimakasih admin informasi
2880	terimakasih informasi
2881	admin terimakasih informasi
2887	terimakasih informasi

Tweet yang telah dikelompokkan kemudian di visualisasikan melalui word cloud. Menunjukkan bahwa kata yang muncul pada cluster pertama adalah “admin”, “informasi” dan “terimakasih”, dimana kata tersebut merupakan kata yang muncul dari 84 tweet di cluster pertama. Beberapa cluster yang terbentuk mempunyai makna yang sama, dimana cluster 2, 6, 9, 10, 15, 16, 24, 26, 29, 31, 33 dan cluster 35 menunjukkan bahwa nasabah ingin meminta customer service atau admin untuk memeriksa, membaca, merespon dan membantu permasalahan nasabah yang ditulis melalui fitur dm atau direct message. Cluster 1, 3, 13, 22, 27 dan cluster 28 menunjukkan bahwa nasabah mengucapkan terimakasih karena admin telah memberikan informasi, edukasi dan tips ataupun telah menyelesaikan masalah nasabah. Cluster 5, 7, 12, 14, 21, 36, 38 menunjukkan keluhan nasabah terkait dengan mobile banking yang error. Cluster 30 menunjukkan sapaan nasabah kepada customer service atau admin perbankan BCA. Sedangkan cluster 4, 8, 11, 17, 20, 23, 32, 34 dan cluster 39 menunjukkan keluhan nasabah terkait dengan transfer rekening, ATM yang tertelan, kartu kredit, dan biaya administrasi.

E. Hasil Clustering DBSCAN

Hasil metode clustering DBSCAN diperlukan untuk mengetahui kriteria dan jumlah cluster optimum yang dihasilkan untuk mengelompokkan tweet pada akun layanan perbankan BRI, Mandiri dan BCA.

Tabel 17. Hasil Cluster pada Layanan Perbankan dengan Metode DBSCAN

Layanan Perbankan	Eps	MinPts	Silhouette Coefficient	Noise	Jumlah Cluster
BRI	0,5	10	0,24065	1.375	33
Mandiri	0,4	15	0,17139	1.370	14
BCA	0,5	10	0,22526	1.440	39

Menunjukkan hasil clustering dari metode DBSCAN dimana pengelompokkan terbaik dari metode DBSCAN adalah dengan melihat kombinasi nilai Eps dan MinPts yang menghasilkan nilai silhouette coefficient yang tertinggi. Dari metode DBSCAN pada layanan perbankan BRI menghasilkan cluster optimum sebanyak 33 cluster dengan kombinasi nilai Eps 0,5 dan MinPts 10 dimana 1.375 tweet tidak bisa dikelompokkan (noise). Pada layanan perbankan Mandiri menghasilkan cluster optimum sebanyak 14 cluster dengan kombinasi nilai Eps 0,4 dan MinPts 15 dimana

1.370 tweet sebagai noise. Pada layanan perbankan BCA menghasilkan cluster optimum sebanyak 39 cluster dengan kombinasi nilai Eps 0,5 dan MinPts 10 dimana 1.440 tweet sebagai noise. Hasil pengelompokan ini menunjukkan hasil yang sama dengan penelitian terdahulu bahwa metode DBSCAN kurang baik untuk mengelompokkan text mining dengan noise yang sangat besar [2].

IV. Kesimpulan

Berdasarkan analisis dan pembahasan yang telah dijelaskan sebelumnya, diperoleh kesimpulan bahwa Perbankan BCA menerima lebih banyak tanggapan atau tweet dari nasabah dibandingkan dengan perbankan BRI dan Mandiri. Informasi yang diperoleh dari ketiga bank tersebut agar customer service bisa melakukan pengecekan pada fitur direct messages dengan kata lain bahwa nasabah menginginkan customer menyelesaikan keluhan mereka dan segera merespon direct messages. Secara garis besar, perbankan perlu melakukan peningkatan pelayanan dari segi kecepatan dalam merespon keluhan customer, keluhan lain berkaitan dengan kartu atm, mobile banking, dan transaksi transfer, dan buku tabungan. Sedangkan dari hasil metode DBSCAN cluster optimal dari perbanka didapatkan dari kombinasi nilai Eps dan MinPts yang menghasilkan nilai silhouette coefficient tertinggi. Nilai silhouette coefficient tertinggi menghasilkan jumlah cluster yang optimal akan tetapi tidak menjamin untuk menghasilkan jumlah noise terendah. Saran bagi peneliti selanjutnya, dapat menggunakan metode pengelompokan lainnya seperti algoritma LVQ. Kelebihan metode ini dapat membentuk suatu pola sehingga dapat menyesuaikan dengan data yang ada[7].

REFERENSI

- [1] Alfarsi. (2017). Data Preprocessing - Konsep Pembelajaran Data Mining. Diakses 23 Desember 2020 dari <https://steemit.com/education/@alfarsi/data-preprocessing-konsep-pembelajaran-data-mining>
- [2] Astutik, D. K. A., Indrasetianingasih, A., dan Fitriani, F. (2022). Penerapan Text Mining pada Analisis Sentimen Pengguna *Twitter* Layanan Transportasi Online Menggunakan Metode Density Based Spatial Clustering of Applications With Noise (DBSCAN) dan K-Menas. *J- Statistika*, Universitas PGRI Adi Buana Surabaya, Vol 15, No 1. 184-189.
- [3] Ariadi, D. & Fithriani, K. (2015). Klasifikasi Berita Indonesia Menggunakan Bayesian Classification dan Support Vector Machine dengan Confix Stripping Stemmer. *Jurnal Sains dan Seni ITS*, Vol. 4, No. 2.
- [4] Arsih, N. (2016). Metode Pengclusteran Berbasis Densitas Menggunakan Algoritma DBSCAN. *Prosiding Statistika*, Vol. 2, Hal. 153-163.
- [5] Detikfinance. (2020). BRI Naik Peringkat di Daftar 1.000 World banks, Nomor 1 di RI. Diakses 17 Desember dari: <https://finance.detik.com/moneter/d-5114772/bri-naik-peringkat-di-daftar-1000-world-banks-nomor-1-di-ri>.
- [6] Feldman, R., & Sanger, J. (2007). *Text Mining Handbook : Advanced Approaches in Analyzing Unstructured Data*. New York: Cambridge University Press
- [7] Fitriani, F. (2016). Pengelompokan data Angin Puting Beliung dengan Menggunakan LVQ. *Wahana* Vol 6 No 1. Universitas PGRI Adi Buana Surabaya.
- [8] Hermanto, E. M.P. (2019). Penerapan Entropy Weighting K-Means (EW K-Means) pada Pengelompokan Kabupaten/Kota di Jawa Timur berdasarkan Indikator Mortalitas Tahun 2014. *Prisma*, Prossiding Seminar Nasional Matematika. Hal 841-847.
- [9] Nasution, M.I.F. & Hidayati, T. (2020). Pengaruh Fitur Chatbot Aisyah (Asisten Interaktif Mandiri Syariah) Terhadap Kualitas Pelayanan Nasabah. *Jurnal BanqueSyar'i*, Vol. 6, Hal 81-88.
- [10] Otoritas Jasa Keuangan. (2020). Laporan Publikasi Bank Umum Konvensional. Diakses 5 Januari dari <https://www.ojk.go.id/id/kanal/perbankan/data-dan-statistik/laporan-keuangan-perbankan/default.aspx>.
- [11] Siregar, A. M., & Puspabhuana, A., (2002). *Data Mining: Pengolahan Data Menjadi Informasi dengan RapidMiner*. Sukoharjo: CV Kekata Group.
- [12] Sulistyawan, S., Hapsery, A., Arifahanum, L. J. A., (2021). Perbandingan Metode Optimasi untuk Pengelompokan Provinsi Berdasarkan Sektor Perikanan di Indonesia (Studi Kasus Dinas Kelautan dan Perikanan Indonesia). *Jurnal Gaussian* Vol 10. No 1., Hal. 76-84.
- [13] Tan, H. P., Plowman, D., & Hancock, P. (2007). Intellectual Capital and Financial Returns of Companies. *Journal of Intellectual Capital*, Vol. 8, Hal 76-91.
- [14] Yusril, A., N. (2020). Implementasi Text Mining untuk Advertising dengan Menggunakan Metode K-Means Clustering pada Data Tweets Gojek Indonesia. *Sistemasi*, Hal. 586-596
- [15] Narsih, A. (2016). Metode Pengclusteran Berbasis Densitas Menggunakan Algoritma DBSCAN. *Prosiding Statistika*, Vol. 2, Hal. 153-163.



© 2023 by the authors. This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License (<http://creativecommons.org/licenses/by-sa/4.0/>).