
Penerapan Text Mining pada Analisis Sentimen Pengguna Twitter Layanan Transportasi Online Menggunakan Metode *Density Based Spatial Clustering of Applications with Noise* (DBSCAN) dan K-Means

Dini Krisnawati Alfiki Astutik⁽¹⁾, Artanti Indrasetianingsih⁽²⁾, Fenny Fitriani⁽³⁾
Program Studi Statistika, Fakultas Sains dan Teknologi, Universitas PGRI Adi Buana Surabaya
Jl. Dukuh Menanggal XII/4, Surabaya 60234
dinialfiki98@gmail.com⁽¹⁾, artanti.indra@unipasby.ac.id⁽²⁾, fenny_f@unipasby.ac.id

ABSTRAK

Transportasi *online* saat ini menjadi populer dan diminati masyarakat di Indonesia dengan transportasi *online* yang banyak digunakan adalah Grab dan Gojek. Meskipun layanan transportasi *online* mendapat respon positif namun terdapat masalah yang dihadapi yaitu banyaknya konsumen yang kecewa dan merasa tidak puas dengan pelayanan yang diberikan. Penelitian ini bertujuan untuk mengetahui pengelompokan tanggapan masyarakat terhadap kedua transportasi *online* tersebut. Tanggapan masyarakat mengenai layanan transportasi *online* didapat dari salah satu media sosial yang banyak digunakan oleh masyarakat Indonesia yaitu twitter. Datapada twitter berupa kumpulan *text* sehingga diperlukan *text mining* untuk menganalisisnya. Salah satu analisis dalam *text mining* adalah *text clustering* sehingga pada penelitian ini menggunakan *text clustering* untuk mengelompokkan pendapat menjadi beberapa kategori. Metode yang digunakan pada *text clustering* adalah metode *Density-Based Spatial Clustering of Applications with Noise* (DBSCAN) dan *K-Means*. DBSCAN adalah sebuah metode yang membentuk *cluster* dari data-data yang saling berdekatan, sedangkan data yang saling berjauhan tidak akan menjadi anggota *cluster* dan biasa disebut sebagai *noise*. *K-Means* adalah teknik *clustering* yang sederhana dan cepat dalam proses *clustering* obyek serta mampu mengelompokkan data dalam jumlah cukup besar. Hasil penelitian menunjukkan bahwa metode DBSCAN dan *K-Means* kurang tepat digunakan pada penelitian ini dalam mengelompokkan *tweet* yang ditujukan kepada layanan transportasi online Gojek dan Grab karena memiliki nilai *silhouette coefficient* kurang dari 0.5 artinya struktur lemah atau *tweet* tanggapan masyarakat kepada layanan transportasi belum berada pada kelompok yang tepat.

Kata kunci : DBSCAN, K-Means, Text Clustering, Transportasi Online

ABSTRACT

Online transportation is now becoming popular and in demand by people in Indonesia with online transportation that is widely used, namely Grab and Gojek. Although online transportation services received a positive response, there were problems faced, namely the number of consumers who were disappointed and dissatisfied with the services provided. This study aims to determine the grouping of public responses to the two online transportation. The public's response to online transportation services is obtained from one of the social media that is widely used by the Indonesian people, namely Twitter. The data on Twitter is in the form of a collection of text, so text mining is needed to analyze it. One of the analyzes in text mining is text clustering, so in this study using text clustering to group opinions into several categories. The methods used in text clustering are *Density-Based Spatial Clustering of Applications with Noise* (DBSCAN) and *K-Means* methods. DBSCAN is a method that forms clusters of data that are close to each other, while data that are far from each other will not be a member of the cluster and are commonly referred to as noise. *K-Means* is a clustering technique that is simple and fast in the object clustering process and is able to group large amounts of data. The results show that the DBSCAN and *K-Means* methods are not appropriate to be used in this study in classifying tweets aimed at online transportation services Gojek and Grab because they have a silhouette coefficient value of less than 0.5 meaning that the structure is weak or the public response tweets to transportation services are not yet in the cluster right.

Keywords : DBSCAN, K-Means, Online Transportation, Text Clustering

1. PENDAHULUAN

Berdasarkan data google temasek, Brain dan Company yang bertajuk economy SEA pada tahun 2019, Indonesia menjadi negara tercepat dan terbesar dalam pertumbuhan ekonomi digital di kawasan Asia Tenggara. Selain *e-commerce*, ekonomi digital Indonesia juga disumbang oleh bisnis transportasi *online*. Fenomena transportasi *online* ini menjadi cepat populer dan diminati oleh masyarakat karena menawarkan inovasi terbaru mengenai transportasi yang digabungkan dengan teknologi komunikasi secara *online*, dan menyediakan berbagai layanan, sehingga memberikan kemudahan bagi masyarakat dalam efisiensi waktu untuk memesan layanan transportasi online dimana saja dan kapan saja dengan hanya menggunakan *smartphone*. Selain itu, transportasi online menjadi salah satu pilihan utama masyarakat karena harga dan kenyamanan.

Hasil survei Asosiasi Penyelenggara Jasa Internet Indonesia (APJI) (2020) periode 2019-kuartal II/2020 mencatat sebanyak 41,9% masyarakat Indonesia menggunakan transportasi online dengan Grab dan Gojek menjadi layanan aplikasi transportasi *online* yang paling sering digunakan oleh masyarakat. Sebanyak 21,3% masyarakat menggunakan Grab dan 19,4% masyarakat menggunakan Gojek. Namun, meskipun transportasi online dapat diterima dengan baik oleh masyarakat di Indonesia, terdapat permasalahan yang dirasakan oleh kedua perusahaan tersebut, yaitu masih banyaknya konsumen yang tidak puas dan kecewa dengan pelayanan yang diberikan. Dominannya pendapat konsumen yang menilai positif pelayanan jasa transportasi online, nampaknya tidak serta merta menghapus kekecewaan konsumen.

Berdasarkan survei yang dilakukan oleh Yayasan Lembaga Konsumen Indonesia (YLKI) pada 5-16 April 2017 diketahui bahwa 59% konsumen transportasi *online* pernah dikecewakan. Banyaknya keluhan konsumen terhadap transportasi *online* menunjukkan bahwa transportasi *online* belum mempunyai standar pelayanan minimal yang jelas dan terukur yang berakibat antara pengemudi satu dengan lainnya tidak sama dalam memberikan pelayanan terhadap konsumennya serta belum mempunyai mekanisme penanganan pengaduan. Tantangan lain yang ada pada konsumen layanan transportasi online saat ini adalah harga. Sebuah penelitian mengenai kesan konsumen tentang ojek *online* juga pernah dilakukan oleh *Research Institute of Economic Development (RISED)*. Hasil penelitian menyatakan bahwa sejumlah 71,12 persen konsumen transportasi *online*

di Indonesia bisa jadi berpaling menggunakan kendaraan pribadi jika tarif transportasi *online* naik (Lingga, Murti Ali. 2019).

Konsumen yang setia merupakan hal yang penting untuk kelangsungan bisnis, jadi bagaimana cara menarik dan mempertahankan loyalitas konsumen. Kepuasan konsumen tentunya merupakan hal penting bagi perusahaan apabila ingin terus bersaing. Konsumen dapat dengan mudah mengekspresikan ulasan, pengaduan, kepuasan dan lainnya dengan mudah melalui berbagai media sosial, salah satunya adalah Twitter. Twitter merupakan media sosial yang populer di Indonesia maupun di berbagai negara lainnya. Menurut data dari Pos dan Informatika Kementerian Komunikasi dan Informatika pada tahun 2020 tercatat sebanyak 19,5 juta dari 245 juta penduduk Indonesia menggunakan twitter. Perusahaan transportasi *online* memiliki akun resmi di Twitter yang menampung *tweet* komentar seperti pelayanan, driver, dan aplikasi, maupun memberikan informasi yang terkini yaitu @gojekindonesia dan @GrabID.

Data dari pendapat masyarakat pada twitter sangat berperan sebagai umpan balik layanan suatu produk atau layanan. Pendapat masyarakat yang ditujukan pada akun resmi transportasi *online* di twitter dapat dikelompokkan menjadi beberapa kategori atau *cluster*. Sehingga perlu dilakukan *text clustering* untuk mempermudah pihak transportasi *online* dalam menanggapi pendapat terkait keluhan, memberikan jawaban dan memberikan informasi kepada masyarakat terkait promosi atau info penting di media sosial twitter. *Text clustering* dapat mengelompokkan kata pada sebuah pendapat yang memiliki kemiripan untuk menjadi beberapa kategori atau *cluster*. Penentuan kategori atau kelompok dari *tweet* yang ditujukan terhadap akun resmi transportasi *online* Gojek dan Grab dalam mengantisipasi banyaknya *tweet*, seperti membuat template tanggapan untuk setiap kategori, dan mengelompokkan *tweet* yang masuk berdasarkan kategori. Sehingga perlu dilakukan penelitian untuk mengetahui tanggapan masyarakat berdasarkan sentimen di media sosial twitter terhadap kedua layanan transportasi tersebut.

Analisis sentimen merupakan metode analisis berbasis komputasi mengenai pendapat, sentimen, evaluasi, sikap, penilaian dan emosi berkenaan dengan suatu topik, produk, layanan, organisasi, individu, ataupun kegiatan tertentu lainnya (Liu, 2010). Sebelum melakukan analisis sentimen, diperlukan praproses teks dengan metode *text mining* untuk mengolah data teks agar siap untuk dianalisis. *Text mining* dapat memberikan solusi dari permasalahan seperti pemrosesan, pengelompokkan, dan menganalisa *unstructured text* dalam jumlah besar.

Terdapat banyak metode *text clustering* yang sering digunakan untuk analisis sentimen namun metode yang sering digunakan dalam *text clustering* adalah metode *Density-Based Spatial Clustering of Applications with Noise (DBSCAN)* dan *K-Means*. Metode DBSCAN adalah algoritma pengelompokan yang didasarkan pada kepadatan (density) data (Tan, dkk., 2007). Metode ini membentuk *cluster* dari data-data yang saling berdekatan, sedangkan data yang saling berjauhan tidak akan menjadi anggota *cluster* (Adinugroho dan Sari, 2018). Sedangkan *K-Means* merupakan metode *clustering* yang memiliki kemampuan untuk mengelompokkan data dalam jumlah yang cukup besar dengan waktu komputasi (Irwansyah dan Faisal, 2015).

Penelitian mengenai analisis sentimen pernah dilakukan oleh Isnawarty, et al (2019) mengenai *text clustering* pada akun twitter layanan ekspedisi JNE, J&T dan POS Indonesia menggunakan DBSCAN dan *K-Means* menghasilkan kesimpulan bahwa DBSCAN menjadi metode lebih baik dari *K-Means* dalam mengelompokkan *tweet* layanan ekspedisi karena memiliki nilai *silhouette coefficient* yang lebih tinggi. Penelitian lain dilakukan oleh Arsih, et al (2016) mengenai metode pengelompokan berbasis densitas menggunakan algoritma DBSCAN terhadap fasilitas kredit nasabah yang menghasilkan kesimpulan bahwa metode DBSCAN bekerja lebih baik dari *K-Means* karena memiliki run time lebih cepat dalam mengelompokkan fasilitas kredit nasabah.

Penelitian ini bertujuan melakukan analisis sentimen mengenai tanggapan masyarakat terhadap layanan transportasi *online* berbasis *text mining* data twitter. Pendapat perlu diidentifikasi dan dikelompokkan menjadi beberapa kategori. Kategori pendapat dapat berupa kepuasan, keluhan, saran, opini, pemberian informasi, dan lain-lain. Pengidentifikasian informasi dari sebuah *text* diperlukan analisis menggunakan *text mining*. Penentuan kategori pengelompokan pada *text mining* dapat dilakukan dengan menggunakan metode *text clustering*. Metode *text clustering* yang digunakan pada penelitian ini adalah metode DBSCAN dan *K-Means*. Dari kedua metode a metode tersebut, akan dipilih metode terbaik yang diharapkan dapat memberikan saran kepada jasa layanan transportasi *online* yaitu Grab dan Gojek dapat mengambil keputusan yang tepat untuk keberlangsungan perusahaan kedepannya dengan mengetahui kategori pendapat yang paling sering muncul dan pengelompokan tanggapan yang diberikan oleh masyarakat dalam mengukur kepuasan dan meningkatkan performa layanan.

2. METODE PENELITIAN

Data yang digunakan pada penelitian ini adalah data yang diambil dari kumpulan *tweet* dari pengguna Twitter di Indonesia dengan akun Twitter yang digunakan dalam analisis adalah akun resmi dari layanan transportasi *online* Grab Indonesia (@GrabID) dan Gojek Indonesia (@gojekindonesia). Data *tweet* diambil pada tanggal 01 Januari hingga 10 Februari 2021 dengan menggunakan Twitter API (*Application Programming Interface*).

Variabel penelitian yang digunakan adalah frekuensi kemunculan kata dasar dari setiap *tweet* yang ditujukan kepada @GrabID dan @gojekindonesia yang telah dilakukan preprocessing dengan skala data rasio.

Langkah analisis yang digunakan pada penelitian ini adalah sebagai berikut:

1. Mengambil data *tweet* dengan menggunakan Twitter API
 - a. Memasukkan *keyword* yang berhubungan dengan akun Twitter @GrabID dan @gojekindonesia
 - b. Menyimpan hasil crawling ke database.
2. Melakukan *text preprocessing* pada data *tweet* layanan transportasi *online* Gojek dan Grab.
 - a. Melakukan *cleaning* data, yaitu menghapus karakter HTML, link URL, *retweet* (RT), username (@username), baris baru, angka, hashtag (#), emoticons, tanda baca, dan *double space*.
 - b. Melakukan *case folding*, yaitu mengubah seluruh teks dengan huruf kecil (non kapital).
 - c. Melakukan *stemming*, yaitu menghilangkan kata imbuhan sehingga didapatkan kata dasar.
 - d. Menghapus kata pada *tweet* yang terdapat pada *stopwords*.
 - e. Melakukan *tokenizing*, yaitu memecah *tweet* menjadi kata per kata.
 - f. Mengubah kata-kata informal menjadi bentuk formal.
 - g. Menghubungkan kumpulan kata yang memiliki makna bila digabungkan.
3. Mengubah data *tweet* ke dalam bentuk frekuensi kemunculan kata dengan menggunakan TF-IDF.
4. Melakukan visualisasi *tweet* dengan *Word Cloud*.
5. Melakukan clustering data

- 5.1. Melakukan clustering data menggunakan metode DBSCAN
 - a. Menentukan parameter *MinPts* dan *Eps*
 - b. Memilih *tweet* *p* secara acak.
 - c. Menandai *tweet* sebagai *core point* dengan berdasarkan nilai *Eps* dan melanjutkan perhitungan jarak antara *core point* dengan *point* yang lain dengan menggunakan rumus jarak *Euclidean*.
 - d. Membuat *cluster* baru dengan menambahkan menambahkan *tweet p* ke dalam *cluster*.
 - e. Identifikasi pada data yang ditandai sebagai *core point* dan melanjutkan sampai semua print telah diproses.
 - f. Jika ada *tweet* yang tidak masuk ke dalam *cluster* manapun akan ditandai sebagai *noise*.
- 5.2. Melakukan clustering data menggunakan metode K-Means.
 - a. Memilih secara acak *k centroid* awal dalam data
 - b. Menentukan jarak setiap kata terhadap pusat *cluster (centroid)*
 - c. Mengelompokkan setiap kata berdasarkan kedekatannya dengan *centroid* terdekat dengan menggunakan *Euclidean distance*.
 - d. Menghitung ulang pusat *cluster (centroid)* baru
 - e. Ulangi langkah (b) hingga (d), sampai anggota yang ada pada tiap *cluster* tidak berubah. Jika jumlah *cluster k* belum diketahui, dapat menggunakan nilai *Variance Ratio Criterion (VRC)* dan metode *Elbow*.
 - f. Membandingkan hasil *clustering* metode *DBSCAN* dan *K-Means* dengan melihat nilai *silhouette coefficient* terbesar untuk menentukan metode *clustering* terbaik yang terpilih
 - g. Melakukan interpretasi dan menarik kesimpulan

3. HASIL DAN PEMBAHASAN

Hasil proses *crawling* data yang telah dilakukan menggunakan *Twitter API* mulai tanggal 01 Januari 2021 hingga 10 Februari 2021 pada akun resmi layanan transportasi *online Gojek* yaitu @GojekIndonesia diperoleh tanggapan pelanggan

sebanyak 2.291 *tweet* dan 2.416 *tweet* yang diperoleh dari akun resmi layanan transportasi *online Grab* yaitu @GrabID.

Text preprocessing merupakan tahapan awal dalam analisis *clustering* yang digunakan untuk mengubah data pada kumpulan *tweet* yang didapatkan dari proses *crawling* menjadi data yang terstruktur sehingga dapat mempermudah analisis selanjutnya. Tahapan yang dilakukan pada *Text preprocessing* yaitu menghapus link, hastag, simbol *retweet (RT)*, *username*, angka, *emoticon*, baris kosong, *punctuation*, spasi berlebih, *case folding*, *stemming*, *stopword*, dan *tokenizing*.

A. Layanan Transportasi Online Gojek

Langkah selanjutnya yang dilakukan setelah melakukan *text preprocessing* adalah melakukan perhitungan frekuensi kemunculan pada masing-masing kata pada setiap *tweet* dan selanjutnya kata tersebut akan dihitung bobot menggunakan *TF-IDF*. Struktur data berdasarkan frekuensi kemunculan kata yang diperoleh setelah dilakukannya *text preprocessing* pada akun layanan transportasi *online Gojek* ditunjukkan pada Tabel 1 sebagai berikut.

Tabel 1. Struktur Data Layanan Transportasi Online Gojek Setelah *Text Preprocessing*

Tweet ke-	Kata						
	akun	...	diskon	...	promo	...	voucher
1	0		1		1		0
2	0		0		0		0
...							
1059	2		0		0		0
...							
2291	0		0		0		0

Jumlah kata yang diperoleh dari proses *text preprocessing* sejumlah 104 kata dasar dan kata-kata tersebut merupakan variabel penelitian yang akan digunakan. Namun, sebelum melakukan proses *text clustering*, perlu dilakukan pembobotan pada kata dasar dari setiap *tweet* dengan menggunakan *Term Frequency – Inverse Document Frequency (TF-IDF)*. Proses perhitungan *TF-IDF* dilakukan dengan menggunakan persamaan (1) dan (2) yang ditunjukkan pada Tabel 2 sebagai berikut.

Tabel 2. Hasil perhitungan *TF-IDF* Layanan Transportasi Online Gojek

Twee t ke-	Kata						
	akun	...	diskon	...	promo	...	vouche r
1	0		0.670 3		0.632 2		0
2	0		0		0		0
...							

1059	0,6589		0		0		0
...							
2291	0		0		0		0

Tabel 2 menunjukkan bahwa kata “diskon” yang terdapat pada *tweet* ke-1 memiliki bobot sebesar 0,6703. Hal ini menunjukkan bahwa kata “diskon” pada *tweet* ke-1 memiliki frekuensi kemunculan yaitu $1 \times 0,6703 = 0,6703$. Begitu pula dengan perhitungan TF-IDF pada kata-kata dasar yang lainnya.

Visualisasi data text menggunakan *Word Cloud* digunakan untuk mengetahui frekuensi kata-kata yang paling sering muncul pada data dengan menggunakan data pada Tabel 2 yang telah boboti menggunakan TF-IDF yang ditunjukkan pada Gambar 1 sebagai berikut.



Gambar 1. Visualisasi *Word Cloud* dari Layanan Transportasi Online Gojek

Gambar 1 menunjukkan bahwa kata yang sering muncul atau memiliki frekuensi tinggi pada tanggapan masyarakat yang ditujukan kepada akun resmi gojek (@GojekIndonesia) adalah kata “dm”, “gojek”, “admin”, “tolong”, “cek”, “order”, “gopay”, “akun”, “gofood”, “bayar”. Sedangkan kata-kata lain yang berukuran kecil artinya kemunculan kata tersebut memiliki frekuensi rendah. Namun, untuk dapat melihat lebih detail apa saja kata-kata yang dan hal yang paling sering dikeluhkan pelanggan kepada akun resmi layanan transportasi online Gojek, data yang telah didapatkan tersebut akan diolah dan dikelompokkan menggunakan metode DBSCAN dan *K-Means*.

Clustering dengan metode DBSCAN pada data *tweet* yang berasal dari akun resmi layanan transportasi online Gojek dilakukan dengan cara melakukan percobaan manual menggunakan

kombinasi nilai *MinPts* dan *Eps* yang berbeda-beda sesuai dengan keinginan peneliti hingga mendapatkan hasil kombinasi *MinPts* dan *Eps* yang terbaik, kombinasi yang digunakan adalah nilai *MinPts* sebesar (5, 15, 25,...95) serta nilai *Eps* antara 0.1 sampai 1. Berikut merupakan hasil 5 kombinasi *MinPts* dan *Eps* yang menghasilkan nilai *silhouette coefficient* tertinggi yang ditunjukkan pada Tabel 3.

Tabel 3. Nilai *Silhouette Coefficient* dan *Noise* metode DBSCAN pada Akun Layanan Transportasi Online Gojek

<i>MinPts</i>	<i>Eps</i>	<i>Silhouette Coefficient</i>	<i>Noise</i>
5	0.6	0.4253	752
5	0.7	0.4082	627
5	0.6	0.4062	858
5	0.5	0.4009	972
5	0.4	0.3958	991

Tabel 4 menunjukkan bahwa hasil *clustering* dari 2.291 *tweet* yang ditujukan kepada akun layanan transportasi gojek menggunakan metode DBSCAN dengan parameter *Eps* sebesar 0,6 dan *MinPts* sebesar 5 menghasilkan nilai *silhouette coefficient* sebesar 0,4253.

Proses *Clustering* menghasilkan jumlah *cluster* sebanyak 69 *cluster* dengan *tweet* yang masuk ke dalam *cluster* sebanyak 1.539 *tweet* dan sisanya sebanyak 752 *tweet* tidak masuk ke dalam *cluster* atau disebut sebagai *noise*.

Hasil *clustering* *tweet* tanggapan pelanggan layanan transportasi gojek menggunakan metode DBSCAN pada *cluster* ke-2 dengan anggota *cluster* sebanyak 19 *tweet*, diperoleh hasil kata yang paling sering muncul yaitu "hapus", "akun", "gojek", "admin", "aplikasi", "gopay", "tolong", "login", "nomor" sehingga diperoleh kesimpulan bahwa kata menunjukkan mengenai keinginan pelanggan untuk menghapus aplikasi gojek dengan penjabarannya ditunjukkan pada Gambar 2 dan Tabel 4 sebagai berikut.



Gambar 2. Visualisasi *Word Cloud* pada *Cluster 2 metode DBSCAN* Layanan Transportasi *Online Gojek*

Clustering menggunakan metode *K-Means*. Dalam metode *K-Means*, perlu menentukan jumlah *cluster (K)* optimum terlebih dahulu dengan menggunakan nilai *Variance Ratio Criterion (VRC)* dan metode *elbow*. Jumlah *cluster* adalah yang memiliki nilai yang paling tinggi. Index yang digunakan untuk menentukan jumlah *cluster* optimum dilakukan dengan *K* mulai 2 sampai dengan 20. Berikut adalah nilai *VRC* pada data layanan transportasi online Gojek yang ditunjukkan pada Tabel 5 dan Gambar 3. Berdasarkan hasil nilai *VRC* pada Tabel 5 dan Gambar 3 menunjukkan jumlah *cluster (K)* yang terbentuk sebanyak 20 dan yang memiliki nilai *VRC* paling tinggi yaitu sebesar 150,386 pada 2 *cluster*. Dimana memberikan nilai *silhouette coefficient* sebesar 0.1771.

Tabel 4. Hasil Clustering pada Cluster ke-2 Metode DBSCAN pada Akun Layanan Transportasi *Online Gojek*

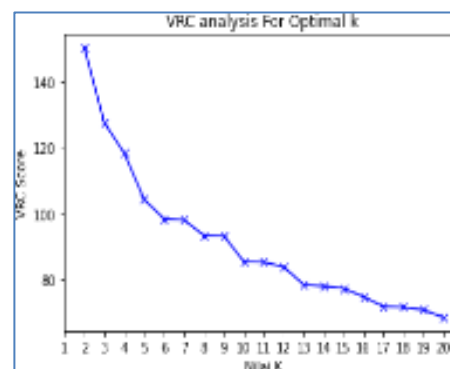
Responden	Tweet	Kata yang paling sering muncul
3	Hallo @gojekindonesia Bagaimana lcara hapus akun gojek y?	"hapus", "akun", "gojek", "admin", "aplikasi", "gopay", "tolong", "login", "nomor"
267	@bapakayammm @BuruhYogyakarta GOJEK sejak dipegang KEVIN memang sadis. driver bonus dihapus, sekarang meres mercha... https://t.co/YgMNORU1Nh	
876	@gojekindonesia Saya mau data2 pribadi sy dihapus/dihanguskan dr database gojek	
880	@gojekindonesia halo, saya ingin mengajukan penghapusan akun gojek saya, karena saya sudah tidak sudi menggunakan a... https://t.co/gecbMvJk0f	
1018	@gojekindonesia alloo gimana cara hapus akun gojek costumer.. please..	
1095	@gojekindonesia hapus permanen	
1096	@gojekindonesia hapus akun gojek bersama gopaynya?	
1332	@gojekindonesia hapus akun min:(klo gitu kan masih ada akunnya bisa login lagi	
	@gojekindonesia kak saya mau	

1499	menghapus akun yg ada ktp saya di gopay plus..karena akunnya hilang
1675	min @gojekindonesia slow respond ah w kan buru mau hapus akun
1715	@bucinnyathehun @gojekindonesia Ya selama tidak di gunakan ^^ Karena kalo dah di hapus maka kamu tidak akan bisa m...

Berikut adalah nilai Metode *Elbow* pada data akun layanan transportasi *online Gojek* yang ditunjukkan pada Tabel 6 dan divisualisaikan pada Gambar 4 sebagai berikut.

Tabel 5. Nilai *VRC* pada layanan transportasi online Gojek.

Cluster (k)	Nilai VRC	Cluster (k)	Nilai VRC
2	150,386	12	83,874
3	127,620	13	78,450
4	118,427	14	78,035
5	104,331	15	77,364
6	98,454	16	74,825
7	98,127	17	71,798
8	93,404	18	71,650
9	93,247	19	70,344
10	85,427	20	68,773
11	85,312		

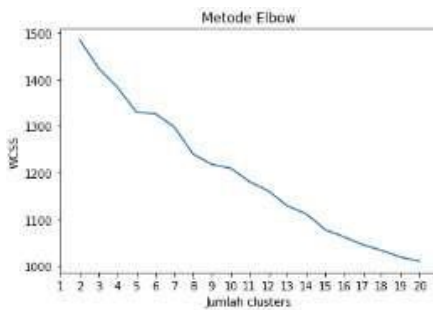


Gambar 3. Nilai *VRC* pada Layanan Transportasi *Online Gojek*

Tabel 6 Nilai Metode *Elbow* pada Layanan Transportasi *Online Gojek*

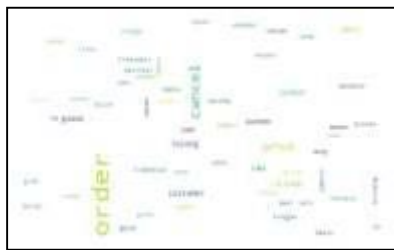
Cluster (k)	Metode Elbow	Cluster (k)	Metode Elbow
2	1485.220	12	1161.421
3	1423.951	13	1129.239

4	1383.177	14	1112.434
5	1329.829	15	1078.576
6	1327.121	16	1063.132
7	1298.649	17	1046.238
8	1240.454	18	1033.773
9	1218.157	19	1019.511
10	1209.940	20	1010.881
11	1180.991		



Gambar 4. Nilai Metode Elbow pada Layanan Transportasi Online Gojek

Berdasarkan hasil nilai metode elbow pada Tabel 6 dan Gambar 4 menunjukkan jumlah cluster (K) yang terbentuk sebanyak 5 karena membentuk siku atau mengalami penurunan nilai SSE. Dimana memberikan nilai *silhouette coefficient* sebesar 0.2138. sehingga dipilih K Optimum yang digunakan adalah 5 karena memiliki nilai *silhouette coefficient* lebih tinggi dibandingkan dengan $K = 2$. Hasil *clustering* pada layanan transportasi online Gojek menggunakan metode *K-Means* dengan visualisasi *word cloud* yang ditunjukkan pada Gambar 5. Hasil *clustering* pada cluster ke-1 layanan transportasi online Gojek menggunakan metode *K-Means* dengan visualisasi *word cloud* yang ditunjukkan pada Gambar 5 sebagai berikut.



Gambar 5 Visualisasi Word Cloud pada Cluster 1 metode *K-Means* Layanan Transportasi Online Gojek

Gambar 5 menunjukkan bahwa hasil *clustering* dari 2.291 tweet yang ditunjukkan

kepada layanan transportasi online Gojek menggunakan metode *K-Means* menghasilkan cluster optimum sebanyak 5 cluster dengan nilai *silhouette coefficient* sebesar 0,2138, dimana cluster 1 dengan anggota cluster sebanyak 112 tweet menunjukkan kata pembatalan pemesanan layanan transportasi online Gojek.

B. Layanan Transportasi Online Grab

Langkah selanjutnya yang dilakukan setelah melakukan *text preprocessing* adalah melakukan perhitungan frekuensi kemunculan pada masing-masing kata pada setiap *tweet* dan selanjutnya kata tersebut akan dihitung bobot menggunakan TF-IDF. Struktur data berdasarkan frekuensi kemunculan kata yang diperoleh setelah dilakukannya *text preprocessing* pada akun layanan transportasi online Grab ditunjukkan pada Tabel 7 sebagai berikut.

Tabel 7. Struktur Data Layanan Transportasi Online Grab Setelah *Text Preprocessing*

Tweet ke-	Kata						
	aku	...	pesan	...	sesuai	...	update
1	1		0		0		0
2	0		1		1		0
...							
1168	0		0		0		1
...							
2416	0		0		0		0

Tabel 7 menunjukkan bahwa jumlah kata yang diperoleh telah berkurang dikarenakan ada beberapa kata yang tidak memiliki makna dalam kalimat sehingga dihapus dan tidak digunakan pada analisis. Jumlah kata yang diperoleh dari proses *text preprocessing* sejumlah 100 kata dasar dan kata-kata tersebut merupakan variabel penelitian yang akan digunakan. Namun, sebelum melakukan proses *text clustering*, perlu dilakukan pembobotan pada kata dasar dari setiap *tweet* dengan menggunakan *Term Frequency – Inverse Document Frequency* (TF-IDF). Hasil perhitungan TF-IDF dapat dilihat pada Tabel 8.

Tabel 8. Hasil perhitungan TF-IDF Layanan Transportasi Online Grab

Tweet ke-	Kata						
	Aku n	...	pesan	...	sesuai	...	update
1	1		0		0		0
2	0		0.6111		0.7915		0

...						
1168	0		0		0	0.6076
...						
2416	0		0		0	0

Tabel 8 menunjukkan bahwa kata “akun” yang terdapat pada *tweet* ke-1 memiliki bobot sebesar 1. Hal ini menunjukkan bahwa kata “akun” pada *tweet* ke-1 memiliki frekuensi kemunculan yaitu $1 \times 1 = 1$. Begitupula dengan perhitungan TF-IDF pada kata-kata dasar yang lainnya.

Visualisasi data text menggunakan *Word Cloud* digunakan untuk mengetahui frekuensi kata-kata yang paling sering muncul pada data dengan menggunakan data pada Tabel 8 yang telah diboboti menggunakan TF-IDF yang ditunjukkan pada Gambar 6 sebagai berikut.



Gambar 6. Visualisasi *Word Cloud* dari Layanan Transportasi *Online Grab*

Gambar 6 menunjukkan bahwa kata yang sering muncul atau memiliki frekuensi tinggi pada tanggapan masyarakat yang ditujukan kepada akun resmi Grab (@GrabID) adalah kata “dm”, “grab”, “kirim”, “admin”, “cek”, “order”, “tolong”, “aplikasi”, “grabfood”, “ovo”. Sedangkan kata-katalain yang berukuran kecil artinya kemunculan kata tersebut memiliki frekuensi rendah.

Namun, untuk dapat melihat lebih detail apa saja kata-kata yang dan hal yang paling sering dikeluhkan pelanggan kepada akun resmi layanan transportasi *online Grab*, data yang telah didapatkan tersebut akan diolah dan dikelompokkan menggunakan metode DBSCAN dan *K-Means*.

Hasil *clustering* menggunakan metode metode *Density-Based Spatial Clustering of Applications with Noise* (DBSCAN) dengan kombinasi yang digunakan adalah nilai *MinPts* sebesar (10, 20, 30,...,60) serta nilai *Eps* antara 0.1 sampai 1. Tabel 9 merupakan hasil kombinasi *MinPts* dan *Eps* yang menghasilkan nilai *silhouette coefficient* tertinggi, yaitu sebesar 0.3419.. Proses *clustering* menghasilkan jumlah *cluster* sebanyak 27 *cluster* dan terdapat 1043 *noise* pada *tweet*.

Tabel 9. Nilai *Silhouette Coefficient* dan *Noise* metode DBSCAN pada Akun Layanan Transportasi *Online Grab*

<i>MinPts</i>	<i>Eps</i>	<i>Silhouette Coefficient</i>	<i>Noise</i>
10	0.6	0.3419	1043
10	0.7	0.3353	936
10	0.5	0.3312	1196
10	0.5	0.3255	1217
10	0.4	0.3192	1232

Berikut adalah visualisasi hasil *clustering* pada cluster ke-1 akun resmi layanan transportasi *online Grab* menggunakan metode DBSCAN yang ditampilkan secara visual menggunakan *word cloud*.

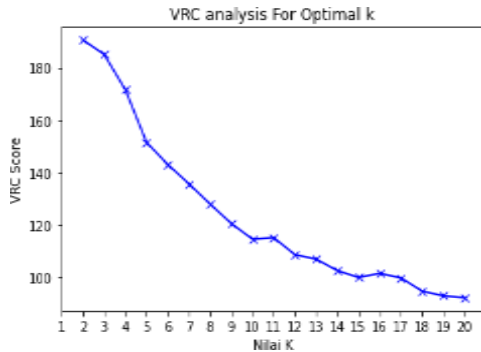


Gambar 7. Visualisasi *Word Cloud* pada Cluster 1 metode DBSCAN Layanan Transportasi *Online Grab*

Hasil *clustering* menggunakan metode *K-Means* pada layanan transportasi *online Grab*. Penentuan jumlah *cluster* optimum berdasarkan dari nilai *Variance Ratio Criterion* (VRC) dan *elbow* yang nilai tertinggi. *Index* yang digunakan untuk penentuan jumlah *cluster* optimum dilakukan dengan *K* mulai dari 2 hingga 20. Berikut adalah nilai VRC pada data akun layanan transportasi *online Grab* yang ditunjukkan pada Tabel 10 dan divisualisaikan pada Gambar 8.

Tabel 10. Nilai VRC pada Layanan Transportasi *Online Grab*

Cluster (k)	Nilai VRC	Cluster (k)	Nilai VRC
2	190.640	12	108.693
3	185.101	13	107.007
4	171.618	14	102.608
5	151.555	15	100,021
6	143.134	16	101.556
7	135.731	17	99.838
8	127.994	18	94.779
9	120.579	19	93.000
10	114,604	20	92.269
11	115,191		



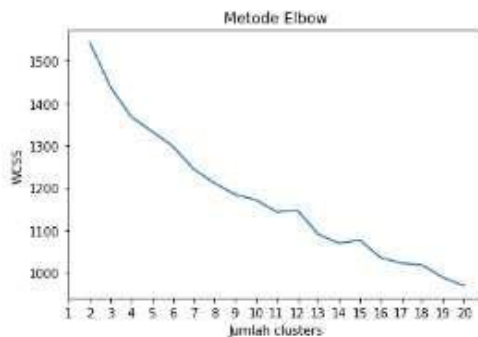
Gambar 9 Nilai VRC pada Layanan Transportasi Online Grab

Berdasarkan hasil nilai VRC pada Tabel 10 dan Gambar 9 menunjukkan jumlah *cluster* (*K*) yang terbentuk sebanyak 20 dan yang memiliki nilai VRC paling tinggi yaitu sebesar 190,640 pada 2 *cluster*. Dimana memberikan nilai *silhouette* sebesar 0.1845.

Berikut adalah nilai Metode Elbow pada data akun layanan transportasi online Grab yang ditunjukkan pada Tabel 11 dan divisualisasikan pada Gambar 10 sebagai berikut.

Tabel 11 Nilai Metode Elbow pada Layanan Transportasi Online Grab

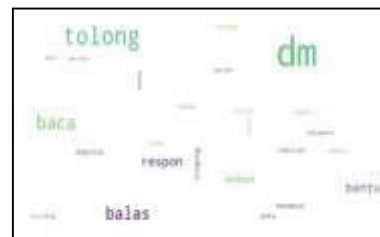
Cluster (k)	Metode Elbow	Cluster (k)	Metode Elbow
2	1544.283	12	1147.376
3	1439.247	13	1090.477
4	1368.063	14	1069.779
5	1333.848	15	1077.132
6	1298.827	16	1035.677
7	1245.134	17	1022.728
8	1211.793	18	1018.164
9	1184.853	19	988.834
10	1172.101	20	969.441
11	1143.816		



Gambar 10 Nilai Metode Elbow pada Layanan Transportasi Online Grab

Berdasarkan hasil nilai metode elbow pada Tabel 11 dan Gambar 10 menunjukkan jumlah *cluster* (*K*) yang terbentuk sebanyak 20 dan yang membentuk siku dan mengalami penurunan nilai SSE pada saat terbentuk 11 *cluster*. Dimana memberikan nilai *silhouette coefficient* sebesar 0.2897. Sehingga dipilih *K* Optimum yang digunakan adalah 11 karena memiliki nilai *silhouette coefficient* lebih tinggi dibandingkan dengan *K* = 2. Hasil *clustering* pada cluster ke-2 layanan transportasi online Grab menggunakan metode *K-Means* dengan visualisasi *word cloud* yang ditunjukkan pada Gambar 11 sebagai berikut.

Berdasarkan hasil nilai VRC pada Tabel 10 dan Gambar 8 menunjukkan jumlah *cluster* (*K*) yang terbentuk sebanyak 20 dan yang memiliki nilai VRC paling tinggi yaitu sebesar 190,640 pada 2 *cluster*. Dimana memberikan nilai *silhouette* sebesar 0.1845.



Gambar 11. Visualisasi *Word Cloud* pada Cluster ke-2 metode *K-Means* Layanan Transportasi Online Grab

Hasil clustering dari 2.416 tweet yang ditujukan kepada layanan transportasi online Grab menggunakan metode *K-Means* menghasilkan sebanyak 11 cluster dengan nilai *Silhouette coefficient* sebesar 0,2897. Gambar 11 merupakan visualisasi *word cloud* pada cluster 2 dengan anggota cluster sebanyak 94 tweet menunjukkan kata permohonan untuk merespon dan cek dan membaca *dm* (*direct message*).

Perbandingan metode DBSCAN dan *K-Means* digunakan untuk mengetahui metode terbaik yang digunakan untuk mengelompokkan tweet pada akun resmi layanan transportasi online Gojek dan Grab. Perbandingan metode clustering ini didasarkan pada hasil nilai *silhouette coefficient* yang diperoleh dari metode DBSCAN dan *K-Means*. Perbandingan dari kedua metode yang digunakan ditunjukkan pada Tabel 12.

Tabel 12. Perbandingan antara Metode DBSCAN dan *K-Means* pada Layanan Transportasi Online Gojek dan Grab

Layanan Transportasi Online	Metode Clustering	Silhouette Coefficient	Jumlah Cluster
Gojek	DBSCAN	0.4253	69
	K-Means	0.1771	2
Grab	DBSCAN	0.3419	27
	K-Means	0.1845	2

Tabel 12 menunjukkan bahwa nilai *silhouette coefficient* pada metode DBSCAN di semua layanan transportasi online Gojek maupun Grab lebih baik dalam mengelompokkan tweet dibandingkan dengan metode K-Means. Hal ini dapat dilihat dari nilai *silhouette coefficient* pada metode DBSCAN lebih besar dibandingkan dengan K-Means. Namun nilai *silhouette coefficient* berada diantara $0,25 < SC < 0,5$ yang berarti termasuk struktur lemah atau pengelompokkan mengenai tanggapan konsumen terhadap layanan transportasi online yang terbentuk memiliki akurasi kurang baik dan belum bisa berpartisipasi tweet ke dalam cluster yang homogen secara tepat.

Metode DBSCAN maupun K-Means kurang tepat diterapkan pada data penelitian ini karena pengelompokkan tweet layanan transportasi grab masih cenderung heterogen atau menyebar secara acak tidak sesuai dengan tujuan dari analisis cluster yaitu mengelompokkan data ke dalam beberapa cluster sedemikian hingga data yang berada di dalam cluster yang sama cenderung mempunyai sifat yang lebih homogen daripada data yang berada di cluster yang berbeda. Namun persaingan diantara kedua perusahaan tersebut tidak jauh berbeda dan sangat kompetitif sehingga perusahaan harus berinovasi secara cepat dan membuat diferensiasi dibandingkan kompetitornya agar dapat meraih pasar yang lebih tinggi.

4. KESIMPULAN DAN SARAN

A. Kesimpulan

Berdasarkan hasil dan pembahasan, maka kesimpulan yang diperoleh sebagai berikut :

1. Berdasarkan hasil crawling data, diketahui bahwa layanan transportasi online Grab memiliki jumlah tweet lebih banyak dibandingkan dengan Gojek. Kata yang paling sering muncul pada tweet yang ditujukan pada layanan transportasi online Gojek adalah “dm”, “Gojek”, “admin”, sedangkan kata yang paling sering muncul pada tweet yang ditujukan pada Grab adalah “dm”, “Grab”, “kirim”.
2. Berdasarkan hasil analisis clustering menggunakan metode DBSCAN pada layanan

transportasi online Gojek memiliki nilai *silhouette coefficient* sebesar 0.4253 dan noise sebanyak 752 tweet dengan 69 cluster yang terbentuk. Sedangkan untuk Grab memiliki nilai *silhouette coefficient* sebesar 0.3419 dan noise sebanyak 1043 tweet dengan 27 cluster yang terbentuk.

3. Berdasarkan hasil analisis clustering menggunakan metode K-Means diperoleh nilai *silhouette coefficient* sebesar 0.1771 pada layanan transportasi online Gojek dan sebesar 0.1845 pada Grab dengan masing-masing terbentuk 2 kluster.
4. Berdasarkan hasil analisis clustering, diketahui bahwa metode DBSCAN merupakan metode terbaik untuk mengelompokkan tweet yang ditujukan kepada akun media sosial twitter layanan transportasi Gojek dan Grab karena memiliki nilai *silhouette coefficient* tertinggi dengan menghasilkan 69 cluster untuk layanan transportasi online Gojek dan 27 cluster pada Grab.

B. Saran

Saran yang dapat diberikan kepada layanan transportasi online Gojek dan Grab yaitu sebagai bahan pertimbangan untuk evaluasi dan meningkatkan performa layanan dalam memberikan respon terhadap pertanyaan dan keluhan pelanggan dengan mengacu pada hasil clustering. Sedangkan saran untuk penelitian selanjutnya, diharapkan agar mendapatkan lebih banyak data dan lebih teliti dalam proses *text preprocessing* karena akan dapat mempengaruhi hasil cluster.

DAFTAR PUSTAKA

- Adinugroho, S. & Sari, Y. A. (2018). Implementasi Data Mining Menggunakan Weka. Malang: UB Press.
- Arsih, Nur., Nusar Hajarisman & Sutawanir Darwis (2016). Metode Pengclusteran Berbasis Densitas Menggunakan Algoritma DBSCAN. *Prosiding Statistika*, Vol. 2, Hal. 153-163.
- Irwansyah, E., & Faisal, M. (2015). *Advanced Clustering : Teori dan Aplikasi*. Yogyakarta: DeePublish.
- Isnawarty, Devi Putri & Irhamah (2019). Text Clustering pada Akun Twitter Layanan Ekspedisi JNE, J&T dan Pos Indonesia Menggunakan Metode Density Based Spatial Clustering of Applications with Noise (DBSCAN) dan K-Means. *Jurnal Sains dan Seni ITS*, Vol. 8, No. 2.

- Lingga, Murti Ali. (2019).
<https://ekonomi.kompas.com/read/2019/02/11/175518026/bila-tarif-ojek-online-naik-konsumen-kembali-ke-kendaraan-pribadi>.
tanggal akses: 20-Des-2020.
- Liu, Bing (2010). Sentiment Analysis and Subjectivity. Chicago : University of Illinois
- Tan, H. P., Plowman, D., & Hancock, P., (2007). Intellectual Capital and Financial Returns of Companies. Journal of Intellectual Capital, Vol. 8, Hal 76-91.