



APLIKASI METODE FUZZY CLUSTERING MEANS UNTUK DATA TRENDING KASUS VAKSIN CORONA PADA JEJARING SOSIAL TWITTER

Damaris Lalang^{1*}, Mariam Lanmay²

^{1,2}Program Studi Matematika FMIPA Universitas Tribuana Kalabahi

Diterima: 10 November 2021 Direvisi: 17 Desember 2021 Diterbitkan : 31 Januari 2022

ABSTRACT

Prevention of the spread of the 2019 corona virus disease by vaccination has become a controversy in a long conversation on Twitter social media. One way to classify these conversations is to use fuzzy clustering means. Fuzzy clustering means is used to group data into several classes with a high degree of similarity, therefore in the study entitled "the application of the fuzzy clustering means method for trending data on corona vaccine cases on the Twitter social network" aims to apply fuzzy clustering means and determine the trending topic of vaccine cases corona. The data is taken and processed using the r studio program with documentation techniques. In the process of implementing the fuzzy clustering means, it was successfully applied by obtaining the optimal value for the validity of the fuzzy silhouette index $x 0.520397533360261$ and being able to classify the data in two clusters and display the words vaccine, covid and vaccination as the most talked about words in the case of corona virus disease 2019 on twitter.

Keywords: fuzzy clustering means, vaccine, corona.

PENDAHULUAN

Perkembangan teknologi komunikasi dan informasi menyebabkan semakin tingginya penggunaan data digital dalam kehidupan sehari-hari. Berbagai bentuk informasi digital berupa data teks diperoleh dengan mudah melalui situs maupun jejaring sosial. Tingginya aktifitas dengan menggunakan data teks dimanfaatkan lebih lanjut untuk melihat kecenderungan dan pola perilaku masyarakat. Analisis pola perilaku dengan menggunakan data teks atau *text mining* banyak dimanfaatkan

produsen untuk melihat produk mana yang cenderung disukai oleh konsumen, atau isu-isu apa yang hangat beredar di masyarakat yang dapat dimanfaatkan untuk merebut perhatian pasar. Dalam jurnal *clustering* berita pada media sosial menggunakan *c-means* diambil data-data teks berita serta mengelompokkan data tersebut dengan menggunakan *c-means* (Rina Candra, 2017). Juga pada jurnal *clustering* pada media sosial dengan *algoritma simple k-means*, melalui penelitian ini dalam pengelompokan akan diperoleh peluang

*Correspondence Address

E-mail: dhamar.ipb14@gmail.com

untuk mengetahui fitur apa yang digemari dan bisa menjadi bahan acuan atau evaluasi para *software development* dalam mengembangkan aplikasi media sosial berdasarkan dengan kelompok kegiatan yang dihasilkan (Rivki, 2018). Dengan melihat isu-isu nasional yang berkembang saat ini bahkan mengancam keberlangsungan di Negara Kesatuan Republik Indonesia lebih menonjol ke wabah Corona virus disease 2019 (covid-19) maka dapat dikelompokkan data tentang *tranding topic* pencegahan covid-19 dengan *Fuzzy Clustering*.

Pada tahun 2020, Negara Indonesia dilanda bencana wabah covid-19 yang juga telah mewabah diseluruh dunia. Pemerintah melakukan berbagai upaya untuk mencegah penyebaran covid-19 yang salah satunya adalah vaksinasi. Pemerintah Indonesia ikut membeli vaksin untuk mencegah penyebaran covid-19 di Negara Indonesia. Penggunaan Vaksin diperuntukkan bagi masyarakat yang sehat jasmani, hal ini menjadi kontraversi di Negara ini, dapat dilihat dalam beberapa postingan negatif yang beredar di media masa seperti Efektivitas vaksin Sinovac belum dapat dipastikan dan Sri Mulyani (Mentri Keuangan) sebut vaksin Covid-19 akan jadi *game changer*. Namun Pemerintah mengambil langkah tepat untuk tetap menggunakannya sebagai cara pencegahan penyebaran Covid-19 yang dipercaya ampuh.

Perkembangan teknologi saat ini, membuat banyak masyarakat yang memilih berpendapat melalui media masa menyangkut dengan penggunaan covid-19. Salah satu media yang digunakan yaitu twitter, menurut data-data survei dari komite penanganan covid-19 dan pemulihan ekonomi nasional (KPCPEN) sekitar 66 % rakyat Indonesia percaya vaksin dan mau divaksinasi covid-19 dan 16 % lainnya tidak mau divaksin atau menolak vaksinasi. Hal ini dapat dilihat pada cuitan-cuitan masyarakat umum di twitter. Akan ada banyak cuitan yang muncul dan untuk mengetahui banyaknya pendapat masyarakat umum mengenai vaksin covid-19 yang bernilai positif, negatif dan netral, serta berapa banyak kelompok kata-kata yang paling banyak digunakan maka data tersebut harus diproses melalui *text mining* dan *fuzzy clustering-means* atau metode lainnya. Pada penelitian ini akan dilakukan pengelompokan data teks untuk mengidentifikasi *tranding topic* pencegahan penyebaran covid-19 pada jejaring *social twitter* dengan metode *fuzzy clustering means*.

Topik-topik yang ada pada jejaring *social twitter* akan digunakan sebagai data pencegahan covid-19 sebagai kata kunci akan mengerucut pada *vaksin covid-19*. Dalam penelitian ini diharapkan dapat menerapkan *fuzzy clustering means* dalam

memperoleh pengelompokan data serta *trending* pada *topic* dengan kata kunci *vaksin covid-19* dapat diidentifikasi. Dengan demikian dapat diperoleh hasil penerapan *fuzzy clustering means* dan informasi *tranding topic vaksin covid-19*.

Pada penelitian ini bertujuan untuk mengetahui implementasi metode *fuzzy clustering means* pengelompokan data teks untuk mengidentifikasi *tranding topic* pencegahan penyebaran covid-19 pada jejaring *social twitter*. Algoritma yang digunakan dalam penelitian ini adalah *Fuzzy Cluster-means* (FCM). FCM menggunakan konsep pendekatan *fuzzy* dimana sebuah objek dapat menjadi anggota dari semua *cluster* yang ada. Derajat keanggotaan tiap-tiap objek dalam *cluster* bernilai antara 0 dan 1. Derajat keanggotaan melambangkan tingkat keberadaan objek didalam *cluster* tertentu (Kusuma, 2004). Dengan menggunakan FCM dapat memperoleh kelas setiap data yang di proses untuk memperoleh informasi *tranding topic* vaksin covid dan implementasi FCM.

METODE PENELITIAN

Penelitian ini dilakukan di *twitter* dengan mengambil data pada jejaring sosial *twitter* dan pengolahan data dengan menggunakan program *R Studio*, penelitian akan dilakukan dengan mengambil data *twitter* selama satu

minggu pada bulan februari 2021. Pada bagian pertama peneliti membuat akun *twitter* dan *Application Programming Interface (API) twitter developer* dan menggunakan kode *token key* dan *secret key* untuk *mengcrawling* data pada *twitter*. Bagian kedua dilakukan pembersihan teks (*teks meaning*) serta pembobotan kata dengan *tf-idf*. Bagian ketiga mengolah data dengan *fuzzy clustering means* dan menarik kesimpulan dari hasil yang didapat. Populasi data adalah 500 *tweets* dan sampel data 140 *tweets* dengan sasaran pada cuitan-cuitan *tweets* pengguna *twitter* tentang *vaksin covid-19*.

Dalam analisis ini menggunakan teknik pengumpulan data melalui dokumentasi sebagai data primernya, dengan mendokumentasikan *twitter* yang telah dipostingkan pada *timeline*. Sedangkan data sekundernya diperoleh dengan *web* dan buku. Peneliti menggunakan teknik pengumpulan data dokumentasi. Dokumentasi adalah suatu cara pengumpulan data yang menghasilkan catatan-catatan penting yang berhubungan dengan masalah yang diteliti, sehingga akan diperoleh data yang lengkap, sah dan bukan berdasarkan perkiraan. Dokumen yang digunakan dapat berupa arsip-arsip hasil dokumentasi *twitter* dari *time line*, foto-foto, dokumen-dokumen seperti Undang-Undang Nomor 14 tahun 2008 dan Instruksi Presiden Nomor 3 Tahun 2003 yang berkaitan dengan

penelitian. Teknik dokumentasi dilakukan untuk melengkapi data yang tidak didapatkan dari proses wawancara. Agar data yang diperoleh peneliti dapat teruji kebenarannya.

Teknik analisis data diadakan setelah semua data terkumpulkan, maka tahap-tahap yang dilakukan adalah sebagai berikut:

1) Proses Teks *Meaning* (*text preprocessing*) yaitu membersihkan data untuk dapat di proses dengan *fuzzy c-means*.

2) Proses *Fuzzy Clustering Means* (*FCM*)

a. Menentukan jumlah *cluster*

Tahap ini akan ditentukan jumlah *cluster*, pangkat untuk matriks partisi, maksimum iterasi, *error* terkecil yang diharapkan, fungsi objektif awal dan iterasi awal.

b. Bangkitkan nilai acak (*random*)

Tahap ini adalah memasukkan data dalam bentuk matriks partisi dan mengubahnya secara random dengan nilai keanggotaan terletak pada interval 0-1.

c. Hitung pusat *cluster*

Tahap ini dilakukan untuk menghitung pusat *cluster* dengan nilai iterasi awal hingga maksimum iterasi jika belum mendapatkan nilai *error* terkecil.

d. Hitung fungsi objektif

Tahap ini sebagai syarat perulangan untuk mendapatkan pusat *cluster* yang tepat.

1. Cek kondisi berhenti

Pada tahap ini akan dilihat kondisi berhenti, jika fungsi objektif lebih kecil atau sama dengan nilai eror terkecil atau iterasi awal lebih besar maksimal iterasi maka akan berhenti. Tetapi jika tidak maka naikkan nilai iterasi untuk melakukan perhitungan pusat cluster ulang hingga mencapai nilai eror terkecil.

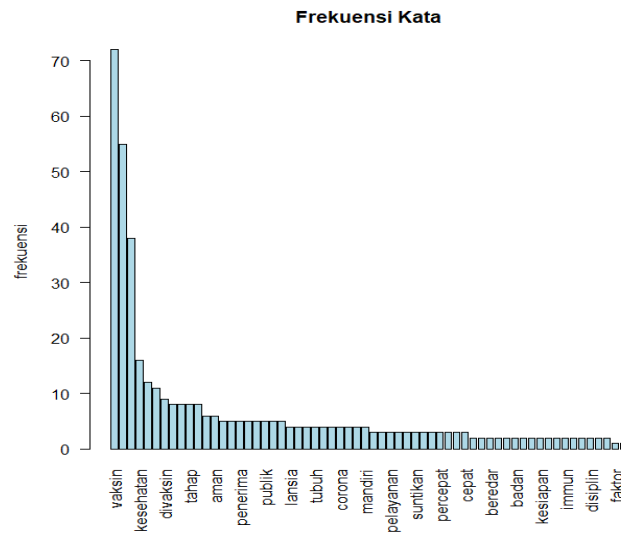
e. Interpretasi dan penarikan kesimpulan

HASIL DAN PEMBAHASAN

Data yang diambil dan diolah dengan *fuzzy clustering means* adalah data berupa teks, yang diambil dari *twitter* dengan kata kunci “*vaksin covid-19*”. Untuk cuitan *tweets* 7 hari (tanggal 5 s.d 11 Februari 2021).

1. Hasil Teks Meaning

Dari 140 data yang dibersihkan menghasilkan 180 *term*. Hasil dari teks meaning berupa *term* dapat ditampilkan dalam bentuk visual grafik berikut.



Gambar 1. Frekuensi Kata

Dari grafik frekuensi kata diatas 1. Pembobotan kata (Tf-Idf) menampilkan kata pertama vaksin dengan frekuensi 72 dan kata ke-61 faktor dengan frekuensi 1. Masih terdapat 119 term dengan frekuensi 1 yang tidak ditampilkan dalam grafik karena akan sulit terbaca dengan jumlah term yang banyak.

Hasil pembobotan tf menggunakan tf biner dimana pada setiap dokumen muncul kata maka diberi bobot satu (1) dan jika tidak ada kata pada dokumen maka di beri bobot nol (0). Hasil pembobotan tf dikali dengan idf menggunakan rstudio akan menghasilkan tfidf sebagai berikut:

No	Atribut												
	wa	wb	wc	wd	we	wf	wg	wh	wi	wj	wk	wl	wm
1	0.000000	0.000000	0.000000	0.000000	2.890372	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
2	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
3	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
4	2.795062	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
5	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000

No	wn	wo	wp	wq	wr	ws	wt	wu	wv	ww	wx	wy	Wz
1	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
2	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	3.113515	0.000000	0.000000	0.000000	0.000000
3	0.000000	0.000000	3.583519	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
4	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
5	0.000000	0.000000	0.000000	0.000000	0.000000	2.5539	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
6	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	2.70805

Gambar 1. Hasil pembobotan kata (*Tf-Idf*)

Pada hasil diatas terdapat 26 atribut dengan 6 dokumen dari 180 dokumen, dapat dilihat pada atribut/kolom yang bernilai 1 keatas terdapat *term* sedangkan pada atribut/kolom yang bernilai 0 tidak ada *term* pada dokumen dan atribut tersebut.

2. Pengolahan Data Dengan Fuzzy

Clustering Means

1. Menentukan banyak cluster

Pada pengolahan data ini hal penting yang dilakukan untuk kelancaran proses data adalah mentukan fungsi objektif awal dan batasan-batasan sebagai berikut :

Nilai Fungsi obyektif awal $(p_0) = 0$

Jumlah cluster $(c) = 2$

2. Bangkitkan nilai acak atau menentukan nilai proporsi data secara *random*

Posisi dan nilai matriks dibangun secara *random*, dimana nilai keanggotaan terletak pada interval 0 sampai dengan 1. Pada posisi awal ini matriks partisi *u* belum akurat. Hasil dari matriks keanggotaan bernilai 0.5 pada cluster 1 dan 2 dari 180 dokumen. Berikut adalah rskrip dan hasilnya.

```
> res.fcm <- fcm(x, centers = 2)
```

```
> #Fuzzy Membership Matrix
```

```
> as.data.frame(res.fcm$u)
```

Cluster 1 Cluster 2

1 0.5 0.5

2 0.5 0.5

3 0.5 0.5

4 0.5 0.5

5 0.5 0.5

6 0.5 0.5

7	0.5	0.5	[1]	1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1
8	0.5	0.5		1 1 1 2 1 1 1 1 1 1 2 1 1 2 1 1 1 1 1 1
9	0.5	0.5		1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 2
10	0.5	0.5	[60]	1 2
				1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
				1 1 2 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
				1 2 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1 2 1 1 1 1
				1 1
			[119]	1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
				1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
				1 1
				1 1 1
			[178]	1 1 1

Hasil atau output data dalam bentuk matriks terdapat 180 dokumen dengan 2 cluster. Pada matriks diatas menampilkan 10 dokumen dari 180 dokumen karena hasilnya sama 0.5 dari 180 dokumen atau term.

3. Hitung pusat cluster atau nilai centroid

Dengan menghitung pusat cluster maka kita akan menemukan titik pusat yang tepat pada cluster tersebut. Pada langkah ini akan mengalokasikan semua term ke cluster pusat terdekat. Hasil output dari rstudio sebagai berikut:

```
> res.fcm2[["cluster"]]
```

Output di atas menunjukkan terbentuknya 2 cluster dari 180 term dengan angka 1 adalah cluster 1 dan angka 2 adalah cluster 2. Namun cluster yang terbentuk tersebut belum optimal, untuk itu dari cluster yang terbentuk ini akan ditentukan nilai pusat clusternya, sebagai berikut:

```
> res.fcm2[["centers"]]
```

	wA	wB	wC	wD	wE	wF
wG	wH					
Cluster 1	0.1708093	0.02884976	0.1383785	0.09954219	0.1605762	0.1133732
	0.126274					
Cluster 2	0.1708093	0.02884976	0.1383785	0.09954219	0.1605762	0.1133732
	0.126274					
	wI	wJ	wK	wL	wM	wN
wO	wP					
Cluster 1	0.1383785	0.0845925	0.126274	0.09954219	0.1133732	0.1383785
	0.09954219					
Cluster 2	0.1383785	0.0845925	0.126274	0.09954219	0.1133732	0.1383785
	0.09954219					
	wQ	wR	wS	wT	wU	wV
wW	wX					

```
Cluster 1 0.1383785 0.1133732 0.1986366 0.0845925 0.1133732 0.1383785 0.09954219
0.1383785
```

```
Cluster 2 0.1383785 0.1133732 0.1986366 0.0845925 0.1133732 0.1383785 0.09954219
0.1383785
```

```
wY wZ
```

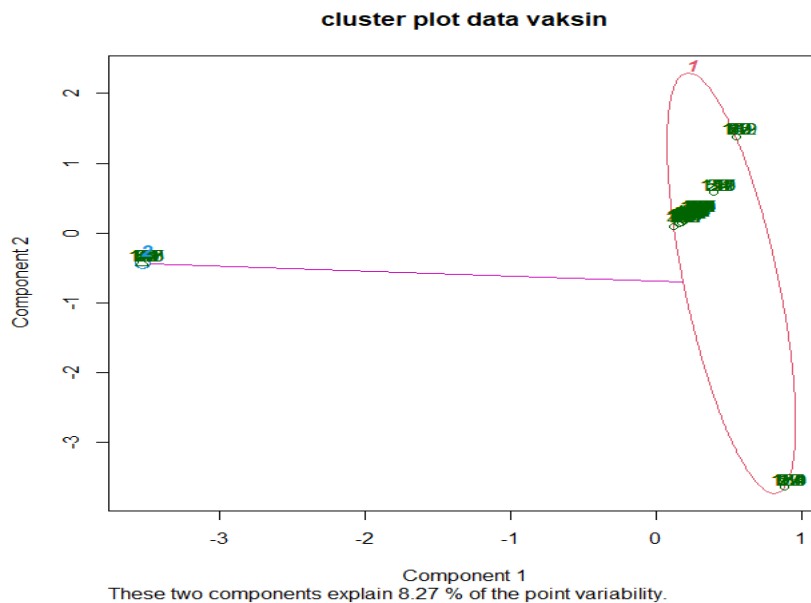
```
Cluster 1 0.1133732 0.1805367
```

```
Cluster 2 0.1133732 0.1805367
```

```
> res.fcm2[["iter"]]
```

```
[1] 20
```

4. Menghitung Fungsi Obyektif Atau Hasil runing dengan visualisasinya
 Menghitung Kembali Nilai Proporsi Untuk dapat terlihat pada gambar berikut.
 Mendapat Pusat Yang Tepat.



Gambar 2. Cluster plot data vaksin

Dari plot data diatas terlihat jelas keberaaan *cluster* 1 dan *cluster* 2 yang memiliki perbedaan cukup jauh.

5. Uji Konvergensi Untuk Melihat Kondisi Berhenti

Perlu dilakukan konvergensi untuk melihat keakuratan suatu data, dengan demikian hasil pengolahan data tersebut akan dikatakan optimal atau tidak optimal. Hasil konvergensi sebagai berikut.


```
> res.fcm4[["clus"]]
 1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29
30 31 32 33
 1  1  1  1  2  1  1  1  1  1  1  1  1  1  1  1  1  1  1  2  1  1  1  1  1  1  2  1
1  2  1  1
34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60
61 62 63 64
 1  1  1  1  1  1  1  1  1  1  2  1  1  1  1  1  1  1  1  1  1  1  2  1  2  1
1  1  1  1
65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91
92 93 94 95
 1  1  1  1  1  1  1  1  1  1  1  1  1  1  2  1  1  2  1  1  1  1  1  1  1  1  1  1
1  1  1  1
96 97 98 99 100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118
119 120 121
 1  1  2  1  1  1  1  1  1  1  1  2  1  1  1  1  1  2  1  1  1  1  1  1
1  1  1
122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144
145 146 147
 1  2  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  2  1  1  1  1  1
1  1  1
148 149 150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170
171 172 173
 1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1
1  1  1
174 175 176 177 178 179 180
 1  1  1  1  1  1  1
```

"Fuzzy Silhouette Index : 0.520397533360261

3. Interpretasi dan Penarikan

Kesimpulan

Hasil pengolahan data vaksin covid-19 dengan *fuzzy clustering means*, pada hasil akhir konvergen dengan menampilkan 2 *cluster* dari 180 *term* sebagai berikut:

Cluster 1: Faktor, vaksin, kedua, vaksinasi, lansia, masyarakat, penggunaan, sinovac, telah, usia, bersama, kerisauan, ketegangan, berhubungan, divaksin, sudah, beredar, covid, informasi, menyelidiki, polisi, video, hingga, Indonesia, pandemi,

kesehatan, penerima, terapkan, mendapatkan, semoga, tahap, pelayanan, utama, meninjau, badan, karantina, menjalani, tidak, kerahkan, kekebalan, tubuh, media, menyiapkan, pemerintah, suntikan, divaksinasi, percuma, sebab, ingat, jadwal, patuhi, mulai, mengumpulkan, soalan-soalan, dinyatakan, mengaku, paranoid, sembuh, mendukung, public, kesiapan, memasok, corona, lindungi, lingkungan, dukungan, mendorong, meningkatkan, penerimaan, protokol, solusi, aturan, penanganan, dukung, percepatan,

pendukung, immune, membantu, ancamanya, nyata, waspada, lawan covid, Aman, kendala, efek, terima, berupaya, memutus, rantai, masker, pakai, wajib, berlanjut, capai, anakanak, terbukti, virus, mandiri, waktu, disiplin, pencegahan, sehat, tensi, efikasi, kemanjuran, khalayak, kontroversi, pengaruh, umum, berkesempatan, cek, kondisi, menerima, mengikuti, lamban, memperketat, satgascovid, kajian, klinik, sukarelawan, terlibat, melibatkan, kegunaan, imunisasi, kementerian, mengikut, perjalanan, konferensi, kerentanan, krusial, perubahan, produk, suhu, kelebihan, berpenyakit, ditunda, memantau, ayo, distribusikan, pastikan, Euphoria, mengabaikan, berikan, manusia, target, cepat, kendalikan, baik, komersialisasi, memastikan, positif, terhindar, cegah, gejala, merasa, penularan, menunggu, apotek, ketersediaan, mengamankan, upaya, batal, dorong, edukasi, gencarkan, ragu.

Cluster 2: darurat, menduga, difabel, tanggulangi, hubungan, lakukan, sosialisasi, terbaik, percepat, penyebaran, terhadap, belum, pengendalian, pemberlakuan.

KESIMPULAN

Dari penelitian yang dilakukan peneliti menarik beberapa kesimpulan sebagai berikut:

1. Metode *Fuzzy Clustering Means* dapat digunakan untuk menganalisis data teks sesuai dengan penjelasan pada hasil penelitian.

2. Dari proses pengolahan data yang dilakukan pada *tranding topic* “pencegahan covid-19” kata yang paling banyak muncul atau dibicarakan sebagai *tranding topic* adalah kata vaksin, covid dan vaksinasi.

DAFTAR PUSTAKA

- Anindya A.P & Zulhanif; 2008. “Jurnal tentang Fuzzy Clustering dalam pengelompokan data teks untuk mengidentifikasi trending topic pariwisata Bali pada jejaring sosial twitter”.
Data.https://www.academia.edu/10378211/Kuliaah_Pengantar_Data_Mining_Penambang_Permata_di_Gunung_Pengetahuan.2003
- Kusuma dewi, Sri; & Purnomo, Hari. (2004). Aplikasi logika fuzzy untuk mendukung keputusan, Edisi Perama Cetakan Pertama. Yogyakarta : Graha Ilmu.
- Rina Candra NS & Sri Eniyati; 2017. “Clustering Berita Pada Media Sosial Menggunakan K-Means ”. (dipublikasi pada tanggal 07 Agustus 2017)
- Rivki Maulana, Iqbal Gibran & Desti Fitriati; 2018. “clustering pada media sosial

dengan algoritma simple k-means,”.
(dipublikasi pada tanggal 01 Desember
2018)

Kusuma dewi, Sri; & Purnomo, Hari. (2004).
Aplikasi logika fuzzy untuk

mendukung keputusan, Edisi Perama
Cetakan Pertama. Yogyakarta : Graha
Ilmu.

