

Analisis Sentimen Komentar Pengguna Twitter Terkait Covid-19 Varian Omicron di Indonesia

Wildan, Moh.Kharis, Moh.Risaldi, Rizaldi Kulap

Prodi Teknik Informatika, Universitas Stimik Adhi Guna, Palu, Indonesia

Info Articles

Keywords:

Sentiment
Analysis; Corona virus
2019; Omicron Variant
; K-Means; TF-IDF; Data
Analytics

Abstrak

Corona virus 2019 berasal dari negeri china, telah menyebar dengan cepat ke penjuru dunia hingga saat ini. Di Indonesia sendiri akhir-akhir ini lagi ramai di perbincangkan di media sosial berupa twitter yakni tweet mengenai Covid-19 varian omicron. Pendapat atau opini dapat dimanfaatkan untuk mencari informasi yang bisa membantu banyak pihak untuk mendukung suatu keputusan. Namun hal tersebut membutuhkan analisis yang tepat. Oleh karena itu, penelitian ini mengusulkan untuk pengenalan sistem analisis sentimen dengan menggunakan metode tf-idf yang dikombinasikan dengan K-means. Analisis Sentimen dengan menggunakan metode tf-idf yang dikombinasikan dengan K-means berhasil membagi data kedalam dua kelompok berbeda, pada cluster satu membahas tentang varian omicron yang terdeteksi. Pada cluster dua membahas tentang protokol kesehatan. Dan mendapatkan nilai Davies Bouldin Index sebesar 0,247. Hasil tersebut dinilai cukup baik sebab semakin mendekati angka nol maka kemiripan data anggota antara cluster semakin baik.

Abstract

Corona virus 2019 originated from China, has spread rapidly throughout the world until now. In Indonesia itself, lately there has been a lot of discussion on social media in the form of Twitter, namely tweets about the Covid-19 variant of the omicron. Opinions or opinions can be used to find information that can help many parties to support a decision. However, this requires proper analysis. Therefore, this study proposes to introduce a sentiment analysis system using the tf-idf method combined with K-means. Sentiment analysis using the tf-idf method combined with K-means succeeded in dividing the data into two different groups, in cluster one discussing the detected omicron variants. In cluster two, it discusses health protocols. And get a Davies Bouldin Index value of 0.247. These results are considered quite good because the closer to zero the better the similarity of member data between clusters.

1. PENDAHULUAN

Corona Virus 2019 berasal dari Negeri China, Kota Wuhan telah menyebar dengan cepat ke penjuru dunia terhitung dari desember 2019 hingga saat ini. Berita tentang covid-19 selalu ramai diperbincangkan dinegeri ini, di Indonesia sendiri diketahui telah masuk beberapa varian covid-19 dan yang akhir-akhir ini lagi ramai di perbincangkan di beberapa media sosial berupa twitter yakni tweet mengenai covid-19 varian omicron yang telah terkonfirmasi masuk di Indonesia pada akhir tahun 2021 lalu, tepatnya pada tanggal 27 bulan november 2021. Dengan mewabahnya pandemic covid-19 varian omicron ini menuai berbagai macam pendapat netizen, mulai dari saran, pujian, hingga kritik terhadap pemerintah dalam upaya penanganannya, bahkan hingga sindiran dan ujaran kebencian. Dengan banyaknya jumlah pengguna twitter yang telah menyampaikan pendapat atau opini dapat dimanfaatkan untuk mencari informasi yang bisa membantu banyak pihak untuk mendukung suatu keputusan. Namun hal tersebut membutuhkan analisis yang tepat, salah satu metode yang dapat di gunakan untuk menganalisa suatu kalimat opini pada twitter adalah analisis sentimen.

Dalam **penelitian Rizqa Luviana Musyarofah, Ema Utami, Suwanto Raharjo. Universitas Amikom Yogyakarta (2020).** Melakukan penelitian dengan judul analisis komentar potensial pada social commerce instagram menggunakan TF-IDF. Tujuan dari penelitian ini adalah menganalisis komentar pada akun toko Instagram dan mengklasifikasikannya ke dalam 2 kelas (potensial dan tidak potensial). Berdasarkan hasil penelitian pada 294 komentar diperoleh hasil akurasi potensi sebesar 80%, nilai presisi 0,76 dan recall 0,94 serta 27% diantaranya adalah komentar tidak potensial. **Penelitian Eko yulian (2018).** Melakukan penelitian dengan mengimplementasi metode K-Means terhadap arsip tweet masyarakat kota bandung dengan tema LGBT, dimana dari lima cluster yang dibentuk diperoleh bahwa kecenderungan cuitan pengguna twitter kota bandung terkait LGBT secara umum masih berhubungan dengan perspektif religi. Kemunculan kata agama yang sangat sering, menyebabkan asosiasi terhadap kata tersebut cukup besar. **Penelitian Try iryanto saputra bersama Rini ariyanti. Universitas Gunadarma Depok (2019).** Melakukan penelitian dengan judul implementasi algoritma k-means clustering pada analisis sentimen keluhan pengguna indosat. Penelitian ini berhasil menampilkan kelompok dari anggota masing-masing cluster yang berbentuk wordcloud ke dalam 3 buah wordcloud berbeda, pada wordcloud cluster 0 anggotanya berbicara tentang jaringan Indosat yang parah, pada wordcloud cluster 1 anggotanya berbicara tentang permintaan perbaikan jaringan sinyal Indosat, pada wordcloud cluster 2 berbicara tentang jaringan sinyal parah Indosat pada daerah Bogor. **Penelitian Andris Faesal, Aziz Muslim, Aditya Hastami Ruger dan Kusrini. Universitas Amikom Yogyakarta (2019).** Melakukan penelitian dengan judul Analisis sentimen pada data tweet pengguna twitter terhadap produk penjualan toko online menggunakan metode k-means. Dengan mengumpulkan sebanyak 1.310 data tweet yang berkaitan dengan penjualan ditokopedia. Penelitian ini berhasil mengelompokkan data tweet ke dalam 3 cluster berbeda yakni kata dengan kemunculan sering, sedang dan jarang didalam data tweet.

Dan dengan pengujian berdasarkan kata tweet diperoleh tingkat akurasi sebesar 92.80%, sedangkan pengujian berdasarkan tweet harian diperoleh tingkat akurasi sebesar 89.80%.

Dalam penelitian ini analisis sentimen digunakan untuk mengetahui pendapat atau kecenderungan opini netizen yang terkandung terhadap suatu permasalahan. Dengan menggunakan metode kombinasi *TF-IDF* dan *K-Means*.

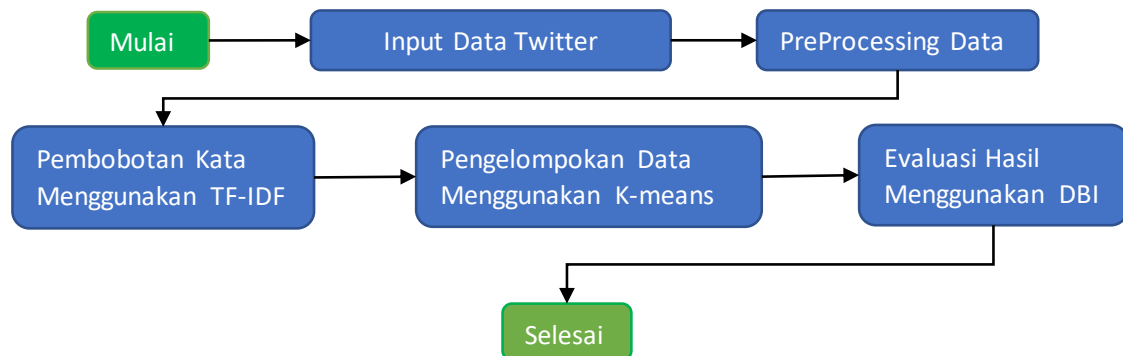
2. METODE

2.1 Pengumpulan Data

Proses *pengumpulan data twitter* ini dilakukan dengan cara meng-*Crawling* otomatis dataset berdasarkan komentar dan cuitan pada twitter yang masi berhubungan dengan kasus covid-19 varian omicron yang masuk diindonesia, dengan menggunakan kata kunci atau hastag #Omicronindonesia. Pengambilan dataset ini menggunakan aplikasi web crawler eksternal yakni netlytic.org Diambil pada tanggal 2 maret 2022 dan berhasil memperoleh sebanyak 131 data. (Data yang diambil pada penelitian ini adalah data terbaru 7 hari terakhir sejak tanggal pengambilan data).

2.2 Alur Sistem

Alur sistem analisis sentimen pengguna twitter terkait covid-19 varian omicron di indonesia dijelaskan pada Gambar 1.

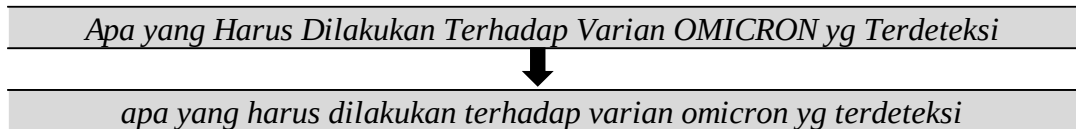


Gambar 1. Alur sistem

2.2.1 PreProcessing

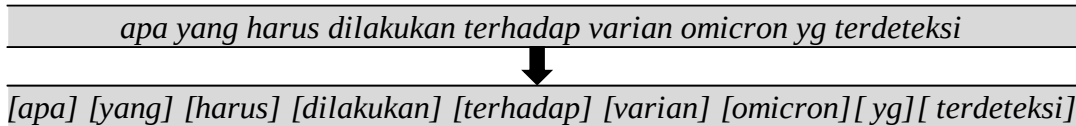
Pada tahap ini data twitter akan dilakukan proses perubahan pada data yang tidak sesuai dan eliminasi pada data yang tidak penting. Sehingga data menjadi lebih terstruktur dan lebih mudah dipahami oleh sistem. Proses preprocessing pada penelitian ini meliputi tahap-tahap sebagai berikut:

- Case Folding. Merupakan proses untuk merubah semua bentuk huruf besar yang ada pada dokumen teks dirubah menjadi huruf kecil.



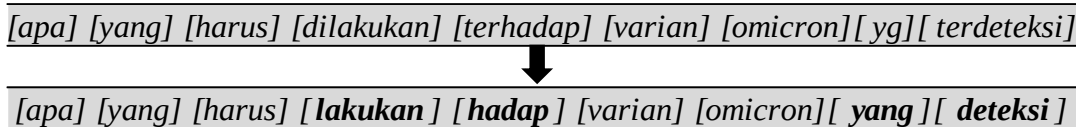
Gambar 2. Case folding

- b) Tokenizing. Dokumen teks yang telah masuk ke proses *tokenizing*, akan berubah menjadi potongan kata yang masing-masing katanya memiliki nilai deskriptif yang berbeda.



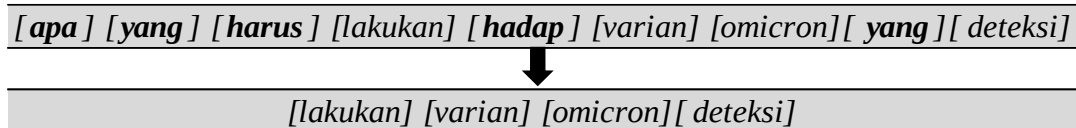
Gambar 3. Tokenizing

- c) Stemming. Pada tahap ini kata dasar pada suatu kalimat akan dirubah menjadi bentuk kata baku bahasa Indonesia yang baik dan benar.



Gambar 4. Stemming

- d) Stopword Removing. Pada tahap ini kata umum yang sering muncul dalam jumlah banyak serta memiliki nilai kecil pada dokumen teks akan dihilangkan.



Gambar 5. Stopword

2.2.2 Term Frequency-Invers Document Frequency (TF-IDF)

Pada tahap ini setiap kata pada dokumen akan diberikan nilai bobot yang berbeda, berdasarkan hasil perhitungan yang dilakukan oleh algoritma *Tf-idf*. Setelah bobot setiap kata berhasil ditemukan. Tahap selanjutnya adalah mencari jumlah bobot setiap kalimat, dengan cara menjumlahkan nilai bobot setiap kata pada dokumen yang ingin yang diketahui nilai bobotnya. Tahap ini bertujuan agar saat proses pengelompokkan pada k-means, data yang dikelompokkan berdasarkan bobot setiap kalimat. Berikut adalah tahapan pembobotan kata dengan algoritma *tf-idf* :

- TF : Jika pada kata **Varian** muncul dua kali dalam satu kalimat maka nilai *tf* untuk kata varian adalah 2. Namun karna kata varian hanya muncul sekali maka nilai *tf* dari kata **varian** ialah 1.
- df : Untuk menentukan suatu nilai *df* yaitu dengan cara menganalisa jumlah dokumen yang mengandung token kata (term). Jika dari 131 dokumen kata varian terdapat pada 55 dokumen berbeda maka nilai *df* dari kata **varian** adalah 55.

IDF : Lanjut ketahap mencari nilai idf yaitu dengan cara membagi jumlah keseluruhan dokumen dengan jumlah dokumen yang mengandung kata (**df**). Menggunakan persamaan (1).

$$IDF = \text{Log} \frac{N}{df} \quad (1)$$

$$IDF = \text{Log} \frac{131}{55}$$

$$= 0,377$$

Weighting

Kata (*term*) : Mencari nilai weighting (**bobot**) suatu kata yaitu dengan cara nilai **tf** kata dikalikan dengan nilai dari **idf** kata. Seperti yang telah dijabarkan pada persamaan (2).

$$W_{t,d} = TF_{t,d} * IDF_t \quad (2)$$

$$W_{t,d} = 1 * 0,377$$

$$W_{t,d} = 0,377$$

Keterangan:

N : Jumlah keseluruhan dokumen dalam korpus

$Df_{(t)}$: Jumlah kemunculan dokumen yang mengandung token kata (*term*)

$TF_{(t,d)}$: Jumlah kemunculan token kata (*term*) pada dokumen *d*

$IDF_{(t)}$: Nilai IDF token *term*

Untuk menemukan bobot tiap kalimat dapat dilihat pada table 1.

Tabel 1. Bobot kalimat

token kata	TF d1	df	IDF	TF-IDF
varian	1	55	0,377	0,377
omicron	1	114	0,06	0,06
lakukan	1	11	1,076	1,076
deteksi	1	22	0,775	0,775
Bobot kalimat =				2,288

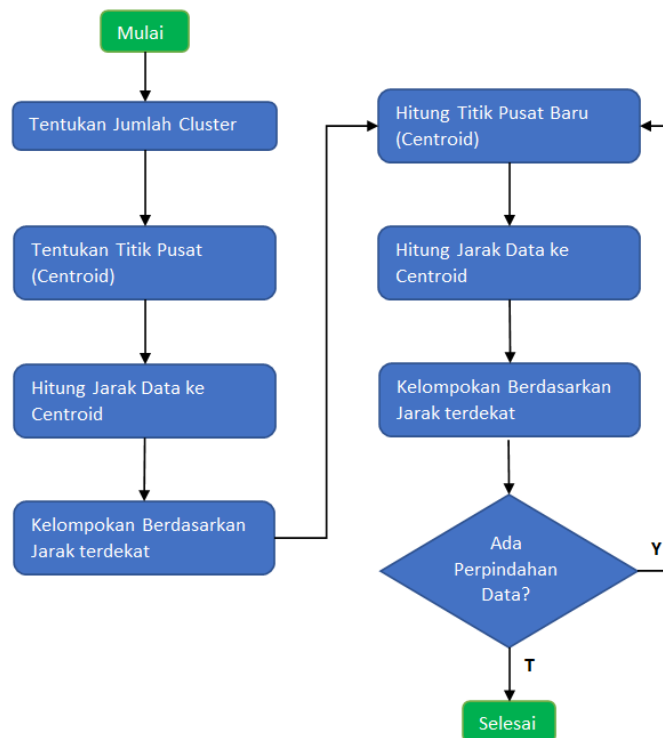
Berikut hasil pembobotan tf-idf untuk 10 dokumen awal dari 131 dokumen komentar:

Tabel 2. TF-IDF

Dokumen	Kalimat	Bobot
d1	varian omicron lakukan deteksi	2,288
d2	omicron inflasi indonesia kendali	5,833
d3	omicron indonesia covid gelombang serang	3,24
d4	varian omicron indonesia landa pandemi covid	3,178
d5	omicron deteksi indonesia ratus siluman waspada gejala	6,626
d6	varian omicron deteksi indonesia covid sakit virus kali rsd wisma atlet kemayoran jakarta pusat dki lakukan vaks	22,235
d7	omicron waspada	1,399
d8	omicron indonesia jakarta total	2,938
d9	varian omicron indonesia capai urutan asia tenggara	8,356
d10	omicron indonesia kalo kena repot bahasa bodoh banget	12,786

2.2.3 K-Means Clustering

Pada tahap ini nilai bobot masing – masing kalimat dari dataset akan dikelompokkan kedalam beberapa kelompok (cluster) berbeda. Berikut adalah alur *K-means Clustering* :



Gambar 6. Alur K-Means Clustering

Tahapan – tahapan untuk implementasi k-means clustering :

- Menentukan jumlah cluster.
- Menentukan data yang akan menjadi titik pusat (*Centroid*).
- Mengelompokan data kedalam cluster sesuai dengan titik centroid terdekat.
- Memperbaharui Centroid.
- Ulangi langkah c). sampai tidak ada lagi anggota data yang berpindah dari cluster satu ke cluster lainnya.

Berikut adalah implementasi k-means clustering dengan menggunakan 10 dokumen awal yang telah melewati tahap preprocessing dan pembobotan kata dapat dilihat di gambar 7.

- Menentukan jumlah cluster. Jumlah cluster yang akan digunakan pada penelitian ini ialah 2
- Menentukan data yang akan menjadi titik pusat (*Centroid*). Data yang digunakan pada centroid awal adalah **2,288** digunakan sebagai *centroid cluster satu (C1)* dan **12,786** sebagai *centroid cluster dua (C2)*
- Mengelompokan data kedalam cluster sesuai dengan titik centroid terdekat. Menggunakan rumus eucliden distance persamaan 3.

$$D_{(x,y)} = \sqrt{\sum_{i=1}^n |x_i - y_i|^2} \quad (3)$$

Setelah data dikelompokan kedalam cluster sesuai dengan titik pusat (*centroid*) terdekat. Diketahui sebanyak 7 anggota data masuk pada cluster 1 dan sebanyak 3 anggota data masuk pada cluster 2 dapat dilihat pada table 3:

Tabel 3. Anggota Cluster

Cluster 1	Cluster 2
2,288	12,786
5,833	22,235
3,24	8,356
3,178	
6,626	
1,399	
2,938	

- Untuk menemukan centroid baru pada masing-masing cluster yakni dengan cara menghitung jarak rata-rata antara anggota cluster lalu dibagi dengan banyak jumlah anggota. Menggunakan persamaan 4.

$$\mu_k = \sum \frac{xq}{nk} \quad (4)$$

Keterangan :

$D(x, y)$ = Jarak antara objek x dan objek y

uk = Nilai centroid baru dari cluster ke- K

xq = Nilai data masing-masing anggota pada cluster ke- K

nk = Banyaknya anggota pada cluster ke- K

2.2.4 Evaluasi Hasil

Tahap terakhir yaitu evaluasi hasil menggunakan **Davies Bouldin Index (DBI)** yang bertujuan untuk menganalisa seberapa baik cluster yang telah dihasilkan oleh K-means. berikut tahapan-tahapan untuk mendapatkan nilai **DBI** :

a) Mencari Nilai **Sum of Square Within (SSW)**

Tahap ini akan menghitung jarak *rata-rata dalam*, anggota data pada masing-masing klaster. Dimana pada tahap ini nilai yang menjadi acuan untuk menghitung SSW adalah *jarak rata-rata dalam anggota data* pada masing-masing cluster pada iterasi terakhir. Menggunakan persamaan 5.

$$SSW_i = \frac{1}{m} \sum_{j=1}^{MJ} d(x_j, c_i) \quad (5)$$

Keterangan :

SSW_i = Rata-rata jarak data dalam setiap klaster.

$d(x_j, c_i)$ = Jarak dalam antar data.

m = Jumlah data dalam setiap klaster

b) Mencari Nilai **Sum of Square Between (SSB)**

Tahap ini akan menghitung jarak antara cluster satu dan cluster dua. Dimana nilai yang jadi acuan atau yang dihitung pada tahap ini ialah nilai centroid akhir pada iterasi terakhir. Menggunakan persamaan 6.

$$SSB_{i,j} = d(x_j - c_i) \quad (6)$$

Keterangan :

$SSB_{i,j}$ = jarak sentroid ke - i dan sentroid ke - j

X_j = nilai sentroid ke- i .

C_i = nilai sentroid ke- j .

c) Mencari Nilai **Ratio**

Pada tahap ini menghitung rasio yakni dengan cara menjumlahkan nilai pada $SSW_{i,j}$ lalu membagikannya dengan nilai $SSB_{i,j}$. Menggunakan persamaan 7.

$$R_{i,j} = \frac{SSW_i + SSW_j}{SSB_{i,j}} \quad (7)$$

Keterangan :

$R_{i,j}$ = rasio antara klaster ke-I dan klaster ke-j

SSW_i = rata-rata jarak dari setiap data ke sentroid ke-i diklaster ke-i.

SSW_j = rata-rata jarak dari setiap data ke sentroid ke-j diklaster ke-j.

$SSB_{i,j}$ = jarak antar sentroid ke-i dan sentroid ke-j.

d) Mencari Nilai **DBI**

Pada tahap ini menghitung nilai dbi dengan cara membagi nilai rasio dengan jumlah klaster. Berikut perhitungannya.

$$DBI = \frac{1}{K} R_{i,j} \quad (8)$$

Ketrangan :

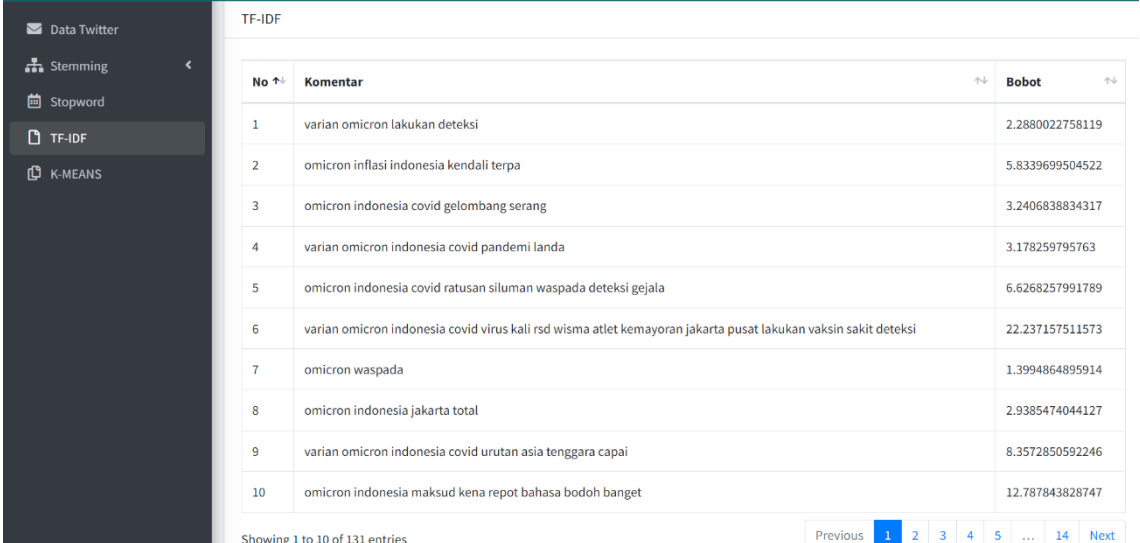
DBI = nilai davies bouldin index.

$R_{i,j}$ = rasio maksimum antar klaster.

k = jumlah klaster.

HASIL DAN PEMBAHASAN

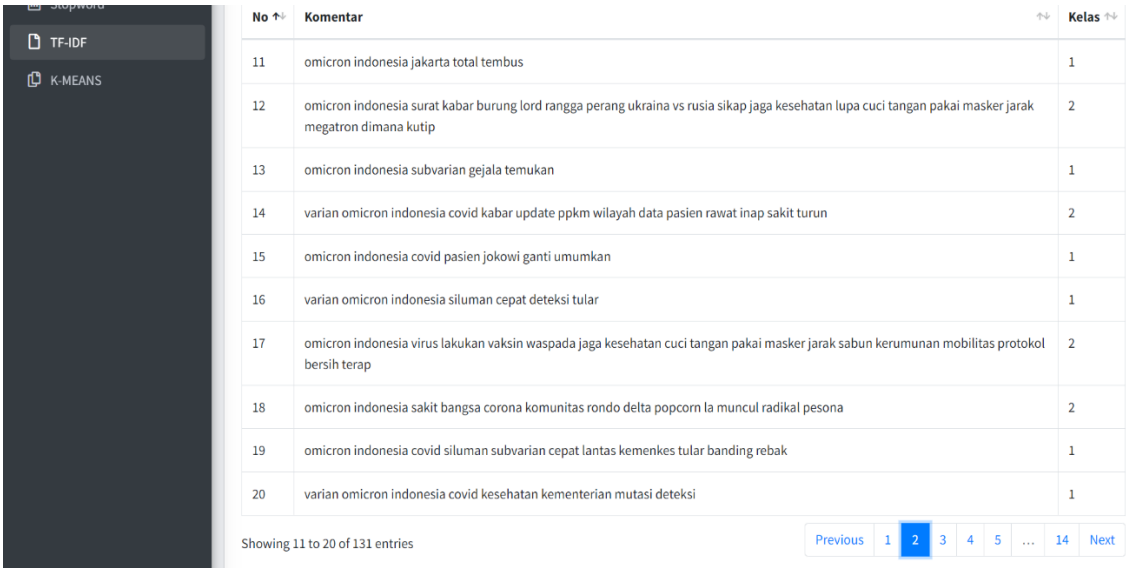
Gambar 7. Merupakan dataset hasil pembootan **tf-idf** yang telah melewati tahap **preprocessing** yang dilakukan oleh sistem.



No	Komentar	Bobot
1	varian omicron lakukan deteksi	2.2880022758119
2	omicron inflasi indonesia kendali terpa	5.8339699504522
3	omicron indonesia covid gelombang serang	3.2406838834317
4	varian omicron indonesia covid pandemi landa	3.178259795763
5	omicron indonesia covid ratusan siluman waspada deteksi gejala	6.6268257991789
6	varian omicron indonesia covid virus kali rsd wisma atlet kemayoran jakarta pusat lakukan vaksin sakit deteksi	22.237157511573
7	omicron waspada	1.3994864895914
8	omicron indonesia jakarta total	2.9385474044127
9	varian omicron indonesia covid urutan asia tenggara capai	8.3572850592246
10	omicron indonesia maksud kena repot bahasa bodoh banget	12.787843828747

Gambar 7. Pembobotan kalimat

Gambar 8. Merupakan hasil implementasi k-means clustering. pada cluster satu mempunyai anggota cluster sebanyak 87 dan cluster dua sebanyak 44 anggota cluster.



No	Komentar	Kelas
11	omicron indonesia jakarta total tembus	1
12	omicron indonesia surat kabar burung lord rangka perang ukraina vs rusia sikap jaga kesehatan lupa cuci tangan pakai masker jarak megatron dimana kutip	2
13	omicron indonesia subvarian gejala temuan	1
14	varian omicron indonesia covid kabar update ppkm wilayah data pasien rawat inap sakit turun	2
15	omicron indonesia covid pasien jokowi ganti umumkan	1
16	varian omicron indonesia siluman cepat deteksi tular	1
17	omicron indonesia virus lakukan vaksin waspada jaga kesehatan cuci tangan pakai masker jarak sabun kerumunan mobilitas protokol bersih terap	2
18	omicron indonesia sakit bangsa corona komunitas rondo delta popcorn la muncul radikal pesona	2
19	omicron indonesia covid siluman subvarian cepat lantas kemenkes tular banding rebak	1
20	varian omicron indonesia covid kesehatan kementerian mutasi deteksi	1

Showing 11 to 20 of 131 entries

Previous 1 2 3 4 5 ... 14 Next

Gambar 8. Hasil clustering

SIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan mendapatkan kesimpulan sebagai berikut: 1).Penerapan metode *tf-idf* yang dikombinasikan dengan *K-Means* dapat melakukan clustering data dalam bentuk teks kedalam dua kelompok berbeda. 2). Analisis sentimen dengan menggunakan metode *tf-idf* dan *k-means* Pada cluster satu kecenderungan cuitan membahas tentang varian omicron yang terdeteksi. Cluster dua cenderung membahas tentang protokol kesehatan 3).Pengujian data mendapatkan nilai *Davies Bouldin Index* sebesar 0,247 yang memiliki arti hasil clustering yang dihasilkan cukup baik. Sebab semakin dekat hasil yang diperoleh dengan angka nol maka kemiripan data anggota antar cluster semakin baik.

Mengingat masi banyak hal-hal yang belum dapat diimplementasikan dari penelitian ini, maka penulis mempertimbangkan saran yaitu melakukan pendekatan dengan metode atau pendekatan yang lain guna mendapatkan hasil penelitian yang lebih baik.

DAFTAR PUSTAKA

- Andris Faesal, A. M. (2019, may). **Sentimen Analisis pada Data Tweet Pengguna Twitter Terhadap Produk Penjualan Toko Online Menggunakan Metode K-Means**. *JURNAL MATRIK*, 207-213. doi.org/10.30812/matrik.v19i2.640
- Han, J.,Kamber, M.,Pei J. 2012 **Data Mining Concept and Techniques**. Wantham: **Elvesier**

- Liu, B. (2012, may). **Sentiment Analysis and Opinion Mining**. 167. Retrieved from <https://doi.org/10.2200/S00416ED1V01Y201204HLT016>
- Musyarofah, R. L. (2020, maret 2). **Analisis Komentar Potensial pada Social Commerce Instagram Menggunakan TF-IDF**. *JURNAL EKSPLORA INFORMATIKA*, 131-139. doi:10.30864/eksplora.v9i2.360.
- Murdick, R.G -1991:27. **Pengertian Sistem Menurut Para Ahli**. (Diakses pada Tanggal 20 November 2021). <http://www.sarjanaku.com/2012/11/pengertian-sistem-menurut-para-ahli.html>
- Robertson, S. (2004, October). **Understanding Inverse Document Frequency**. *Journal of Documentation*, 60. doi:10.1108/00220410410560582
- Supomo, P. V. (2016). **Pemograman Web Dengan Menggunakan Php dan Codeigniter**. *Deepublish*, 1-4.
- Suyanto, (2017). **Data Mining Untuk Klasifikasi dan Klasterisasi Data**. Bandung : Informatika Bandung.
- Tan, P.N., Steinbach, M., Kumar, V. (2006) *Introduction to Data Mining*. Boston : Pearson Education.
- Try Iryanto Saputra, R. A. (2019, Desember 3). **Implementasi Algoritma K-Means Clustering Pada Analisis Sentimen Keluhan Pengguna Indosat**. *Jurnal Ilmiah Informatika Komputer*, 24, 191-198. doi:10.35760/ik.2019.v24i3.2361.
- Yulian, E. (2018, Mei). **Text Mining dengan K-Means Clustering pada Tema LGBT dalam Arsip Tweet Masyarakat Kota Bandung**. *Jurnal Matematika Mantik*, 53-58. doi:10.15642/mantik.2018.4.1.53-58