

Tweet Analysis of Mental Illness Using K-Means Clustering and Support Vector Machine

Analisis *Tweet* Gangguan Kesehatan Mental Menggunakan K-Means *Clustering* dan Support Vector Machine

Kartikadyota Kusumaningtyas¹, Muhammad Habibi², Irmma Dwijayanti³, Retno Sumiyarini⁴

^{1,2} Informatika, Universitas Jenderal Achmad Yani Yogyakarta, Indonesia

³ Sistem Informasi, Universitas Jenderal Achmad Yani Yogyakarta, Indonesia

⁴ Keperawatan, Universitas Jenderal Achmad Yani Yogyakarta, Indonesia

¹kartikadyota@gmail.com @gmail.com, ^{2*}muhammadhabibi17@gmail.com,

³irmmadwijayanti@gmail.com, ⁴retno.sumiyarini@gmail.com

Informasi Artikel

Received: June 2023

Revised: August 2023

Accepted: September 2023

Published: October 2023

Abstract

Purpose: Social media, particularly Twitter, provides a venue for individuals to share their thoughts. The public's perception of mental illnesses is often debated on Twitter. So yet, no evaluation of community tweets connected to data on mental health conditions has been performed. The purpose of this study is to examine tweets linked to mental illnesses in Indonesia in order to identify the themes of conversation and the polarity trends of these tweets.

Design/methodology/approach: To address this issue, the K-Means Clustering algorithm is utilized to aggregate tweet data that is used to find themes of conversation. The emotion polarity value of each cluster result was then determined using the Support Vector Machine (SVM) approach.

Findings/results: This study generated five topic clusters based on tweets about mental illness. While sentiment analysis revealed that all clusters had more negative sentiment classes than positive. Cluster 4 and Cluster 5 had the highest number of negative sentiment values. These clusters emphasize the necessity of consulting with psychiatrists and psychologists if people have mental health disorders, as well as financing for mental health disorder treatment through BPJS Kesehatan services.

Originality/value/state of the art: The analysis was done in two stages: data grouping to find themes of conversation using K-Means clustering and SVM to look for positive and negative polarity values associated to twitter data about mental illness.

Keywords: Sentiment Analysis;
Clustering; Mental Illness
Kata kunci: Analisis sentimen;
Clustering; Gangguan Kesehatan
Mental

Abstrak

Tujuan: Media sosial khususnya Twitter memberikan wadah untuk mengungkapkan opini kepada penggunanya. Opini masyarakat terkait gangguan kesehatan mental banyak dibahas di Twitter. Evaluasi *tweet* masyarakat terkait data gangguan kesehatan mental selama ini belum dilakukan. Penelitian ini bertujuan untuk menganalisis *tweet* terkait gangguan kesehatan mental di Indonesia, sehingga dapat diketahui topik pembahasan dan kecenderungan polaritas yang dimiliki oleh *tweet-tweet* tersebut. .

Perancangan/metode/pendekatan: Untuk mengatasi permasalahan tersebut, digunakanlah algoritma K-Means *Clustering* untuk melakukan pengelompokan data *tweet* yang digunakan untuk mencari topik pembahasan. Setelah itu dilakukan proses analisis sentimen dengan menggunakan algoritma Support Vector Machine (SVM) untuk mencari nilai polaritas sentimen dari masing-masing hasil *cluster*.

Hasil: Penelitian ini menghasilkan 5 *cluster* pembicaraan terkait *tweet* gangguan kesehatan mental. Sedangkan hasil analisis sentimen didapatkan semua *cluster* memiliki jumlah kelas sentimen negatif lebih banyak dibandingkan dengan positif. *Cluster* yang memiliki jumlah nilai sentiment negatif paling signifikan adalah *cluster* 4 dan *cluster* 5. *Cluster* tersebut membahas mengenai pentingnya konseling dengan psikiater dan psikolog jika mengalami gangguan kesehatan mental serta mengenai pembiayaan pengobatan gangguan kesehatan mental melalui layanan BPJS Kesehatan. .

Keaslian/state of the art: Analisis dilakukan dengan dua tahapan yaitu pengelompokan data untuk mencari topik pembahasan menggunakan K-Means *clustering* dan SVM untuk mencari nilai polaritas positif dan negatif terkait data *tweet* gangguan kesehatan mental.

1. Pendahuluan

Berdasarkan data dari Kementerian Kesehatan dalam beberapa tahun terakhir presentase masyarakat Indonesia yang mengalami gangguan kesehatan mental meningkat. Riset Kesehatan Dasar (Riskesdas) yang dilakukan Kementerian Kesehatan tahun 2018 menunjukkan prevalensi Rumah Tangga dengan anggota menderita gangguan jiwa skizofrenia meningkat dari 1,7 permil menjadi 7 permil di tahun 2018 [1]. Gangguan mental emosional pada penduduk usia dibawah 15 tahun, juga naik dari 6,1% atau sekitar 12 juta penduduk menjadi 9,8% atau sekitar 20 juta penduduk. Kondisi ini diperburuk dengan adanya COVID-19, masalah gangguan kesehatan jiwa dilaporkan meningkat sebesar 64,3% baik karena menderita penyakit COVID-19 maupun masalah sosial ekonomi sebagai dampak dari pandemi [2].

Twitter merupakan salah satu media sosial yang memiliki pengguna paling banyak dibandingkan media sosial yang lain. Berdasarkan data dari katadata, pengguna Twitter Indonesia masuk urutan kelima jumlah pengguna Twitter terbanyak di dunia, yaitu sebanyak 24 juta pengguna pada awal tahun 2023 [3]. Media sosial khususnya Twitter memberikan wadah untuk mengungkapkan opini kepada penggunanya [4]. Opini masyarakat terkait gangguan kesehatan mental banyak dibahas di Twitter. Evaluasi *tweet* masyarakat terkait data gangguan kesehatan mental selama ini belum dilakukan. Untuk mengetahui hasil evaluasi tersebut, maka *tweet* harus dianalisis sehingga dapat diketahui *tweet* tersebut memiliki topik pembahasan serta memiliki kecenderungan polaritas apa saja.

Tujuan dari penelitian ini adalah untuk menganalisis *tweet* terkait gangguan kesehatan mental di Indonesia, sehingga dapat diketahui topik pembahasan dan kecenderungan polaritas yang dimiliki oleh *tweet-tweet* tersebut [5]. *Tweet* mengenai gangguan kesehatan mental di Indonesia akan dilakukan proses *clustering* untuk mengetahui pembahasan apa saja yang diperbincangkan dalam topik tersebut. Selain itu, hasil analisis *clustering* yang sudah didapatkan akan dilakukan analisis sentimen berupa polaritas positif, dan negatif. Sehingga penelitian ini akan menghasilkan hasil analisis perbincangan terkait gangguan kesehatan mental masyarakat berdasarkan *clustering* dan dapat dilihat nilai polaritas sentimen dari hasil *cluster* tersebut [6].

Penelitian yang menggunakan data gangguan kesehatan mental sudah banyak dilakukan, diantaranya adalah penelitian mendeteksi depresi dan penyakit mental di media sosial [7], penelitian ini menghasilkan metode deteksi otomatis dapat membantu mengidentifikasi individu yang depresi atau berisiko melalui pemantauan pasif berskala besar di media sosial. Penelitian lain adalah prediksi awal dan perjalanan penyakit mental dengan data Twitter [5], penelitian ini menyarankan pendekatan prediktif berbasis data untuk skrining dini dan deteksi penyakit mental. Selain itu penelitian untuk menilai depresi dan penyakit mental di media sosial menggunakan Twitter [6], hasil dari penelitian ini mengeksplorasi penyebab yang menimbulkan kemurungan dan keresahan dari gangguan kesehatan mental.

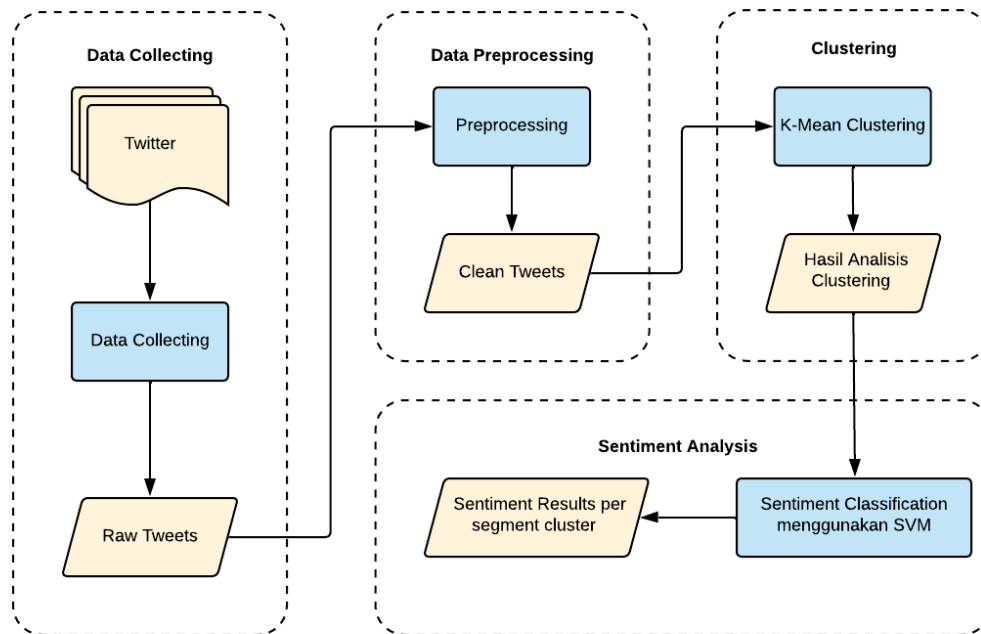
Penelitian ini menggunakan algoritma K-Means sebagai solusi untuk mengelompokkan suatu data *tweet*. Algoritma K-Means banyak digunakan dalam penelitian terkait dengan pengklasteran data, diantaranya adalah penelitian terkait klasterisasi user karakteristik berdasarkan *hashtag* pada *platform* Instagram [8], serta klasterisasi tipe pembelajar sebagai parameter evaluasi kualitas pendidikan di perguruan tinggi [9]. K-Mean *Clustering* merupakan salah satu algoritma yang digunakan untuk melakukan *object mining* yang bersifat tanpa latihan (*unsupervised learning*). K-Means *Clustering* digunakan untuk melakukan analisis *cluster* non hirarki yang berusaha untuk mempartisi objek yang ada ke dalam satu atau lebih *cluster* atau kelompok objek berdasarkan karakteristiknya [10]. Sehingga objek yang mempunyai karakteristik yang sama dikelompokkan ke dalam satu *cluster* yang sama dan objek yang mempunyai karakteristik yang berbeda akan dikelompokkan ke dalam *cluster* lain. Tujuan dari *clustering* atau pengelompokkan adalah untuk meminimalkan *objective function* yang di atur dalam proses *clustering*, pada dasarnya, digunakan untuk meminimalkan variasi di dalam satu *cluster* dan memaksimalkan variasi antar *cluster* [11].

Penelitian terkait klasifikasi teks sudah banyak dilakukan, penelitian ini menggunakan algoritma Support Vector Machine (SVM) untuk melakukan klasifikasi sentimen. Adapun penelitian terkait klasifikasi teks diantaranya adalah penelitian terkait komentar mahasiswa [12], klasifikasi judul skripsi [13], klasifikasi jurnal berdasarkan abstrak [14]. Data media sosial juga

banyak digunakan untuk penelitian, seperti pengelompokan *followers* influencer berdasarkan like dan komen [15], analisis *trend* percakapan terkait *marketplace* [16]. Analisis sentimen merupakan salah satu implementasi dari klasifikasi teks. *Sentiment analysis* atau opinion mining merupakan bidang studi yang menganalisis tentang pendapat orang, sentimen, evaluasi, penilaian, sikap, dan emosi terhadap entitas produk, jasa organisasi, individu, isu, kejadian, topik, dan atribut-atributnya [17]. sedangkan pengertian lain menyebutkan bahwa *sentiment analysis* merupakan analisis subjectivitas, *opinion mining* dan ekstraksi penilaian dengan beberapa hubungan yang terkait dengan komputasi afektif yaitu *computer recognition* dan ekspresi emosi.

2. Metode

Penelitian ini memiliki beberapa tahapan yang dilakukan, diantaranya adalah pengambilan data *tweet* dari Twitter, kemudian dilakukan *preprocessing*, setelah itu dilakukan *clustering* data menggunakan K-Means untuk mendapatkan topik pembahasan. Langkah selanjutnya dilakukan proses analisis sentimen untuk mengetahui nilai sentimen setiap *cluster* data yang didapatkan. Tahapan penelitian dapat dilihat pada Gambar 1.



Gambar 1. Tahapan Penelitian

2.1. Pengumpulan Data

Data pada penelitian ini diperoleh dari hasil ekstraksi data *tweet* Twitter dengan menggunakan metode *web data extraction*, yaitu sistem perangkat lunak yang secara otomatis dan berulang mengekstrak data dari suatu halaman *web* dengan konten yang berubah dan mengirimkan data yang diekstraksi ke suatu database atau aplikasi lain [18]. Data yang diambil adalah *tweet* ataupun *re-tweet* bahasa Indonesia dengan menggunakan kata kunci “gangguan kesehatan

mental”. Data yang digunakan dalam penelitian ini dikumpulkan dari tanggal 1 Januari 2021 sampai dengan 31 Maret 2023 dengan perolehan data sebanyak 4.337 *tweet*.

2.2. Preprocessing Data

Tahap selanjutnya yaitu *preprocessing*, *preprocessing* adalah suatu cara yang dilakukan sebelum melakukan proses *data mining* sehingga menghasilkan data yang memiliki arti yang lebih sederhana [19]. Tahap *preprocessing* terbagi beberapa tahapan proses antara lain, proses pertama *case folding* yaitu proses untuk mengubah huruf kapital menjadi huruf kecil atau *lowercase*. Proses yang kedua tokenisasi yaitu membagi *tweet* menjadi bentuk token yang dipisahkan dengan spasi. Proses yang ketiga *stopword removal* yaitu menghilangkan kata penghubung serta kata yang tidak memiliki suatu makna. Dan proses yang terakhir *stemming* yaitu mengubah suatu kata berimbuhan menjadi kata dasar. Setelah data *tweet* bersih, tahapan berikutnya adalah pengelompokan data *tweet* menggunakan K-Means *Clustering*. Fitur yang digunakan baik untuk K-Means dan Support Vector Machine adalah menggunakan *Term Frequency-Inverse Document Frequency* (TD-IDF).

2.3. Clustering

Tahapan berikutnya adalah melakukan pengelompokan atau *clustering* data. *Clustering* merupakan penggunaan teknik *data mining* dimana sekelompok objek yang sama digabungkan bersama untuk membentuk *cluster*, *cluster* ini berbeda dari objek di *cluster* lain [20]. Pada tahap ini digunakan K-Means untuk mengklasterisasi *tweet*, sehingga diketahui pembahasan apa saja yang diperbincangkan. Adapun perhitungan K-Means diperoleh dari persamaan (1) berikut:

$$d(x_j, c_j) = \sqrt{\sum_{j=1}^n (x_j + c_j)^2} \quad (1)$$

Keterangan:

d = jarak

n = banyaknya objek

j = (dimulai dari 1 sampai n)

x_j = *feature* objek ke j terhadap x

c_j = *centroid feature* ke j

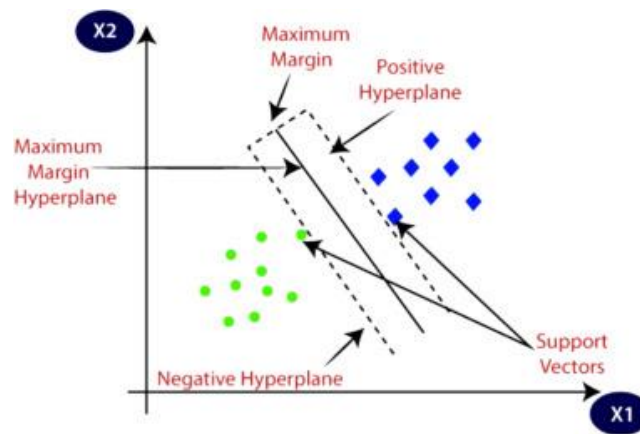
Secara umum langkah algoritma K-Means [9] adalah:

- a. Menentukan banyaknya *cluster* (k)
- b. Menentukan *centroid*
- c. Analisis apakah *centroid* berubah berdasarkan *means* setiap *feature* objek (data)?
 - 1) [ya], ubah *centroid* menjadi *centroid* baru
 - 2) [tidak], selesai
- d. Hitung *distance space* (jarak kedekatan) objek dengan *centroid*
- e. Kelompokan objek berdasarkan kedekatan objek dengan *centroid*

Langkah selanjutnya adalah melakukan analisis sentimen serta membuat demografi terhadap data tersebut.

2.4. Sentiment Analysis

Tahap selanjutnya yaitu melakukan klasifikasi tingkat polaritas setiap *tweet* dengan menggunakan analisis sentimen. Pada tahap ini, digunakan pendekatan metode *supervised learning* dengan memanfaatkan data latih untuk membentuk model analisis sentimen. Algoritma yang digunakan untuk analisis sentimen adalah Support Vector Machine (SVM) [21]. Dalam algoritma SVM, titik *vector* ekstrim yang disebut Support Vectors dipilih yang membantu dalam membuat *hyperplane* yang sesuai [22]. Algoritma SVM biasa digunakan dalam deteksi wajah, klasifikasi beberapa gambar, pengklasifikasian teks, dll. Diilustrasikan pada Gambar. 2 dimana dua kategori kontras diklasifikasikan menggunakan algoritma SVM. Misalnya, seseorang menemukan suatu *tweet* positif yang memiliki ciri-ciri yang mirip dengan *tweet* negatif. Oleh karena itu, algoritma SVM cukup menguntungkan untuk membuat model yang dapat mengidentifikasi dengan tepat apakah itu *tweet* dengan polaritas positif atau *tweet* dengan polaritas negatif.

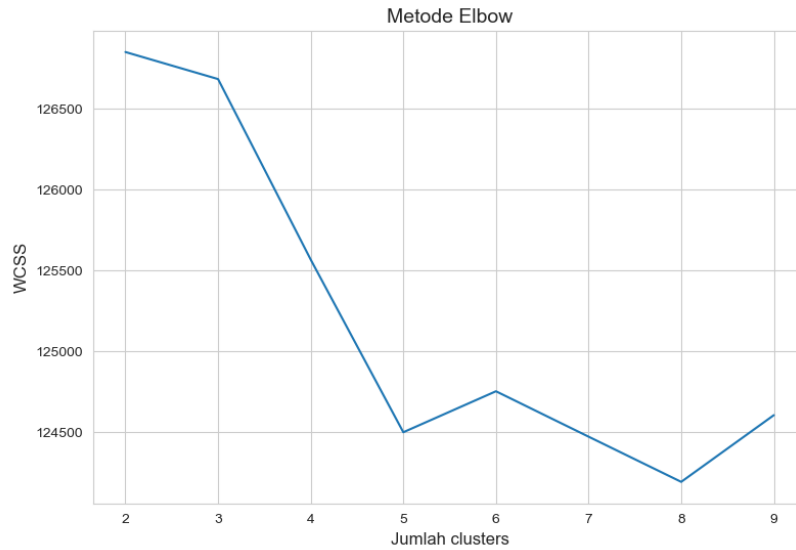


Gambar 2. Ilustrasi SVM [23]

3. Hasil dan Pembahasan

3.1. Clustering

Pada tahapan ini dilakukan membangun segmen *cluster* dengan rentang yang telah ditentukan, kemudian melakukan evaluasi terhadap jumlah *cluster* yang paling sesuai untuk analisis yang lebih lanjut. Penelitian ini menggunakan metode elbow untuk menentukan jumlah *cluster* terbaik. Metode elbow adalah teknik yang diterapkan untuk menghasilkan data dalam menentukan jumlah *cluster* yang optimal dengan cara memperhatikan persentase hasil perbandingan antara jumlah *cluster* yang akan membentuk sudut pada titik tertentu [24]. Gambar 3 menunjukkan hasil jumlah *cluster* terbaik menggunakan metode elbow.



Gambar 3. Hasil metode elbow

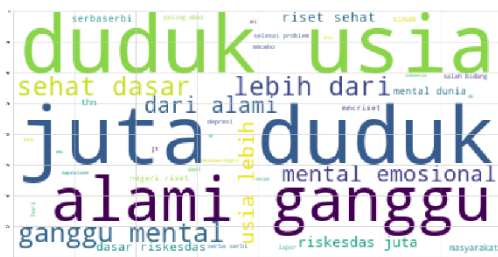
Berdasarkan Gambar 3, terlihat bahwa jumlah *cluster* optimal adalah 5, karena titik siku pertama muncul pada nilai 5. Oleh karena itu, analisis berikutnya akan menggunakan 5 *cluster*. Kami telah membuat model visualisasi *wordcloud* untuk analisis konten, yaitu *wordcloud* untuk setiap *cluster*. Visualisasi *wordcloud* akan memberikan gambaran tentang kata-kata yang sering digunakan dalam percakapan tertentu. Ini berarti, kita dapat melihat kata-kata yang paling signifikan dalam membangun segmen *cluster* tertentu dari model yang dibangun. Gambar 4 menunjukkan *wordcloud* untuk data 5 *cluster*.



Cluster 1



Cluster 2



Cluster 3



Cluster 4



Cluster 5

Gambar 4. Wordcloud masing-masing cluster

Berdasarkan hasil *wordcloud* pada Gambar 4, maka kami membuat analisis pembahasan untuk masing-masing *cluster*. Analisis pembahasan dibuat berdasarkan hubungan kata-kata yang sering muncul pada masing-masing *cluster*, sehingga didapatkan pembahasan pada setiap *cluster*. Untuk menganalisis *cluster* data Twitter, penting untuk tidak hanya mengandalkan frekuensi kata kunci untuk menentukan topik apa yang ada di dalam setiap *cluster*. Pendekatan yang lebih mendalam juga harus mempertimbangkan seluruh konteks atau isi *tweet*. Metode ini memungkinkan kita untuk mendapatkan pemahaman yang lebih luas tentang pembicaraan yang terjadi, bukan hanya berdasarkan kata-kata yang sering muncul, tetapi juga tentang aspek-aspek yang lebih dalam dari percakapan yang terjadi dalam *cluster* tersebut. Metode ini memungkinkan kita untuk memperoleh pemahaman yang lebih akurat dan mendalam tentang tren dan pendapat yang ada di data Twitter. Hasil klastering *tweet* dapat dilihat pada Tabel 1.

Tabel 1. Hasil klastering *tweet*

No	<i>Tweet</i> bersih	Cluster
1	halusinasi bisa sebab beberapa faktor ganggu sehat mental pengaruh obatobatan bius lihat dengar buruk ganggu pola tidur kurang oksigen racun gas monoksida sakit berat	1
2	kondisi sering pada milik kondisi sehat mental kurang seperti alami depresi ganggu cemas orang yang alami clinomania cenderung takut laku aktivitas seharihari	1
3	ganggu sehat mental timbul dengan tidur saat sama kurang tidur insomnia jadi faktor kontribusi muncul buruk gejala sehat mental	1
4	indonesia darurat sehat mental dasar survei laku dokter spesialis dokter jiwa indonesia naik ganggu jiwa pandemi covid yuk simak live react with us	2
5	sehat jiwa dan masyarakat dunia buah prioritas utama pilih tema ini kaca tingkat ganggu cemas lebih persen masa pandemi layan sehat mental	2
6	tanggal oktober kemarin hari sehat mental dunia sebab pandemi covid who tahun benar dunia juang hadap beban global ganggu jiwa	2
7	ribu warga bandung alami ganggu jiwa akibat pandemi survei warga kota bandung lebih dari warga alami sehat mental	3
8	dasar riset sehat dasar jumlah remaja derita ganggu emosi mental persen pada tingkat persen	3
9	psikolog klinis dewasa annisa prasetyo ningrum universitas indonesia jumlah ganggu sehat mental risiko muncul tengah pandemi kerap kait cemas depresi	3
10	alami fase quarter life crisis rasa cemas apakah cemas yang alami masuk ganggu sehat mental jelas psikolog	4
11	tulis orang sintas ganggu cemas butterfly hug urai detail proses sadar cari bantu masalah sehat mental ia hadap dengan self-diagnosis dengan bantu medis psikolog	4
12	periksa laku tes diagnosis ganggu cemas psikiatri dokter ahli sehat mental akan laku evaluasi psikologis pasien	4

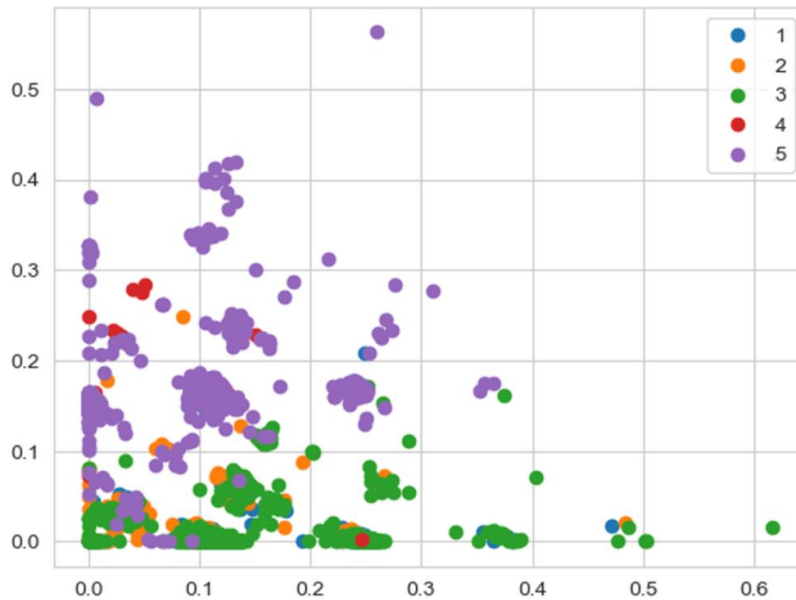
No	Tweet bersih	Cluster
13	ganggu sehat mental tu gak orang gila aja loh cemas depresi masuk loh langsung aja periksa pake bpjs biaya gratis	5
14	kasus ganggu sehat mental tak atas faskes maka dokter beri rujuk rumah sakit maupun rumah sakit jiwa milik kompetensi jiwa telah sama bpjs sehat	5
15	bpjs tanggung biaya konsultasi obatobatan agam ganggu jiwa depresi bipolar skizofrenia gratis program jknkis jamin layan masalah sehat mental sesuai prosedur tidak membedakan layan serta jknkis	5

Analisis topik pembahasan pada Twitter bukan hanya menghitung berapa kali kata kunci muncul; metode yang hanya berfokus pada frekuensi dapat menghasilkan kata-kata yang sering muncul dan tampak dominan, tetapi kata-kata tersebut seringkali tumpang tindih dan memberikan gambaran yang kurang jelas. Oleh karena itu, isi atau konteks *tweet* juga digunakan sebagai referensi penting untuk menentukan topik yang akan dibahas dalam setiap *cluster*. Metode ini memastikan bahwa topik yang ditemukan mencerminkan topik sebenarnya dari diskusi, bukan hanya kata-kata yang sering diulang. Hasil analisis setiap *cluster* dapat dilihat pada Tabel 2.

Tabel 2. Analisis pembahasan pada setiap *cluster*

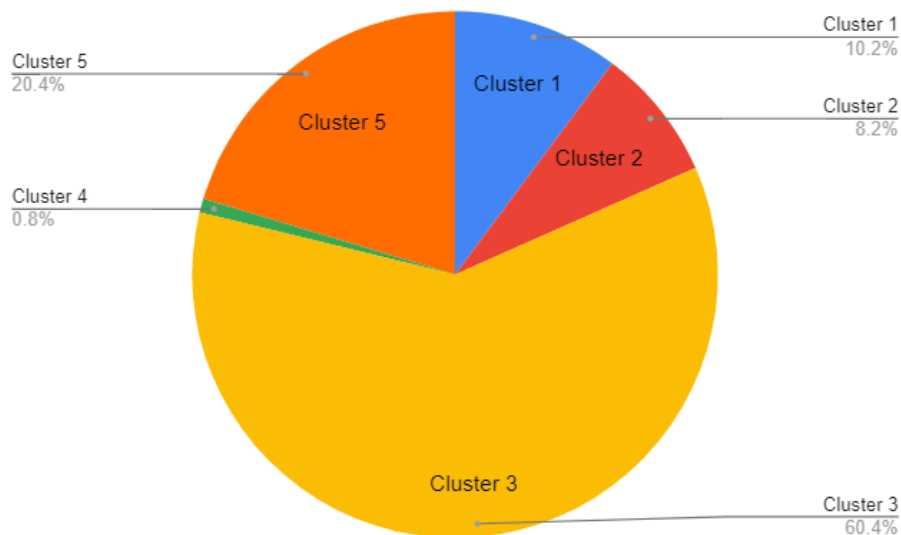
Cluster	Kata Kunci	Analisis Pembahasan
Cluster 1	Sehat, ganggu, mental, cemas, sakit, alami, stress, depresi, gejala, sebab	Gejala gangguan kesehatan mental
Cluster 2	Sehat, mental, pengaruh, stigma, masyarakat, ganggu, kondisi, kena, hidup, pandemi	Pengaruh pandemi Covid-19 terhadap gangguan kesehatan mental
Cluster 3	Alami, ganggu, mental, penduduk, juta, sehat dasar, lebih dari, emosional, usia lebih, riskesmas	Jumlah penderita gangguan kesehatan mental di Indonesia
Cluster 4	Gangu, mental, sehat jiwa, sakit, psikiater, psikolog, obat, tingkat, sadar, fisik	Pentingnya melakukan konsultasi dengan psikiater dan psikolog terkait kesehatan mental
Cluster 5	Sehat, mental, bpjs, biaya, layanan, biaya bpjs, keluarga, sakit, pasien, obat	Pembiayaan berobat gangguan kesehatan mental menggunakan layanan BPJS Kesehatan

Hasil Analisis pembahasan untuk setiap cluster dapat dilihat bahwa *Cluster 1* membahas mengenai gejala-gejala gangguan kesehatan mental, kemudian *cluster 2* membahas mengenai pengaruh pandemi Covid-19 terhadap meningkatnya penderita gangguan kesehatan mental. Adapun *cluster 3* membahas mengenai jumlah penderita gangguan kesehatan mental di Indonesia. *Cluster 4* membahas mengenai pentingnya melakukan konseling terhadap psikiater dan psikolog untuk preventif dan pengobatan gangguan kesehatan mental. Sedangkan *cluster 5* membahas mengenai pembiayaan pengobatan gangguan kesehatan mental melalui layanan Badan Penyelenggaraan Jaminan Sosial (BPJS) Kesehatan. Gambar 5 merupakan visualisasi 5 *cluster* yang dihasilkan oleh K-means. Pada gambar tersebut dapat dilihat persebaran data *cluster* tidak merata, ada *cluster* yang memiliki jumlah data *tweet* yang cukup banyak dan ada juga *cluster* yang memiliki jumlah data *cluster* yang cukup sedikit. Hal tersebut tergantung dari jumlah pembahasan mengenai topik yang diperbincangkan terkait gangguan kesehatan mental.



Gambar 5. Arsitektur sistem *social media intelligence*

Pada Gambar 6 dapat dilihat bahwa *cluster* yang memiliki jumlah data *tweet* terbanyak adalah *cluster* 3 yaitu membahas mengenai jumlah penderita gangguan kesehatan mental di Indonesia. Kemudian disusul dengan *cluster* 5 yaitu membahas mengenai pembiayaan pengobatan gangguan kesehatan mental menggunakan layanan BPJS kesehatan. Pembahasan dengan jumlah data *tweet* paling sedikit adalah *cluster* 4, yaitu mengenai pentingnya konsultasi dengan psikiater dan psikolog jika mengalami gangguan kesehatan mental. *Cluster* 4 memiliki jumlah *tweet* yang paling rendah dikarenakan, masih kurang sadarnya masyarakat terhadap pentingnya konsultasi dengan psikolog atau psikiater jika mengalami gangguan kesehatan mental. Presentase jumlah data *tweet* masing-masing *cluster* dapat dilihat pada Gambar 6.



Gambar 6. Persentase jumlah *tweet* tiap *cluster*

3.2. Analisis Sentimen

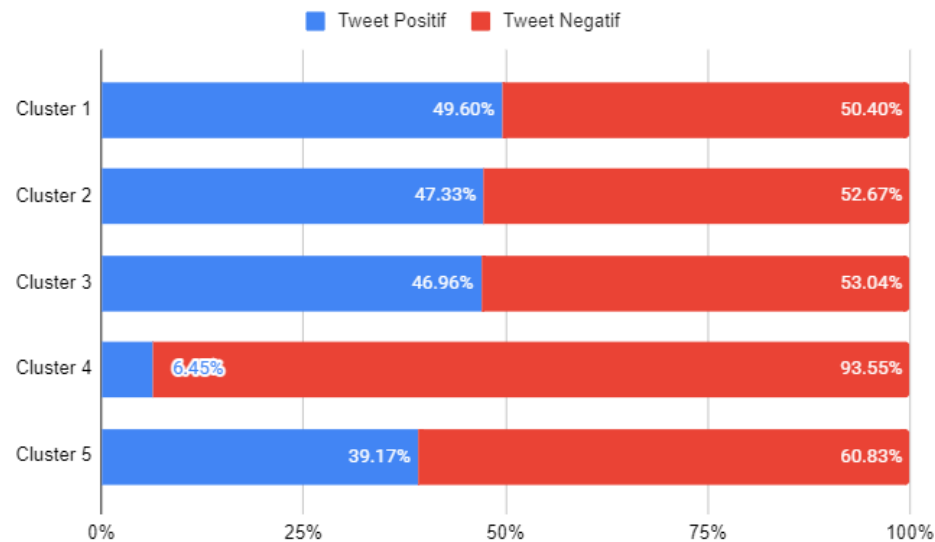
Setelah topik pembahasan pada setiap *cluster* terkait *tweet* gangguan kesehatan mental teridentifikasi, tahap selanjutnya adalah melakukan analisis sentimen pada setiap *cluster*. Analisis sentimen yang dilakukan mencakup klasifikasi sentimen menjadi dua kategori, yaitu positif dan negatif. Dengan demikian, informasi mengenai pandangan masyarakat terhadap *tweet* gangguan kesehatan mental dalam setiap *cluster* dapat diketahui dengan jelas. Proses analisis sentimen dilakukan dengan memanfaatkan algoritma Support Vector machine (SVM) yang telah melalui pelatihan dengan tingkat akurasi 81,73%. Data latih yang digunakan telah diberi label secara manual pada 1.442 *tweet*, terdiri dari 720 *tweet* positif dan 722 *tweet* negatif. Sementara itu, terdapat 350 data *tweet* yang digunakan untuk pengujian. Hasil analisis sentimen menunjukkan kelas sentimen pada tiap *cluster*, sebagaimana yang terlihat pada Tabel 3.

Tabel 3. Hasil analisis sentiment tiap *cluster*

Cluster	Jumlah Tweet	
	Sentimen Positif	Sentimen Negatif
Cluster 1	186	189
Cluster 2	142	158
Cluster 3	1043	1178
Cluster 4	2	29
Cluster 5	293	455

Gambar 7 merupakan grafik hasil analisis sentimen tiap *cluster*. Berdasarkan hasil analisis sentimen tersebut, dapat diketahui jumlah *tweet* kelas sentimen positif atau negatif pada tiap *cluster*. Dari semua *cluster* didominasi lebih banyak *tweet* dengan kelas sentiment negatif dari pada sentiment positif, walaupun dengan selisih yang tidak terlalu banyak. *Cluster* yang memiliki jumlah *tweet* negatif paling signifikan dibanding dengan *tweet* positif yaitu *cluster* 4 dan *cluster* 5. *Cluster* 4 membahas mengenai pentingnya konseling dengan psikiater dan psikolog jika mengalami gangguan kesehatan mental. Hal ini dapat diartikan banyak masyarakat yang belum memiliki kesadaran untuk memelihara kesehatan mental [25], padahal hal tersebut sangat penting. Peran psikiater dan psikolog sama-sama membantu mengatasi masalah mental. Serta masih tabunya anggapan masyarakat kalau orang konsultasi ke psikolog dan psikiater itu dianggap memiliki gangguan kejiwaan.

Pada *Cluster* 5, membahas mengenai pembiayaan pengobatan gangguan kesehatan mental melalui layanan BPJS Kesehatan. pada *cluster* tersebut juga memiliki jumlah selisih yang signifikan antara *tweet* negatif dengan *tweet* positif. Hal tersebut mengindikasikan bahwa sebagian masyarakat masih belum mengetahui jika psikolog atau psikiater di rumah sakit atau puskesmas biaya perawatan kesehatan di keduanya masih menjadi tanggungan BPJS Kesehatan. BPJS Kesehatan juga bisa digunakan untuk rehabilitasi medis dan konseling dengan psikolog [26].



Gambar 7. Hasil persentase analisis sentiment tiap cluster

4. Kesimpulan

Tujuan utama penelitian ini adalah melakukan analisis data *tweet* terkait gangguan kesehatan mental menggunakan menggunakan K-Means *clustering* dan Support Vector Machine. Untuk melakukan analisis tersebut, penelitian ini melakukan dua tahapan analisis yaitu pengelompokan data untuk mencari topik pembahasan menggunakan K-Means *clustering* dan analisis sentimen untuk mencari nilai polaritas positif dan negatif. (1) Hasil *clustering*, didapatkan 5 cluster pembicaraan terkait gangguan kesehatan mental. Sedangkan hasil analisis sentimen, semua cluster memiliki jumlah kelas sentimen negatif yang lebih banyak dibandingkan dengan positif. (2) Cluster yang memiliki jumlah nilai sentiment negatif paling signifikan adalah cluster 4 dan cluster 5. Cluster tersebut membahas mengenai pentingnya konseling dengan psikiater dan psikolog jika mengalami gangguan kesehatan mental serta mengenai pembiayaan pengobatan gangguan kesehatan mental melalui layanan BPJS Kesehatan. Hasil analisis yang didapatkan diharapkan dapat dijadikan bahan evaluasi agar seluruh lapisan masyarakat lebih peduli terhadap isu gangguan kesehatan mental yang sekarang sedang marak dialami oleh masyarakat, baik generasi tua maupun generasi muda.

Daftar Pustaka

- [1] D. H. Jayani, "Persebaran Prevalensi Skizofrenia/Psikosis di Indonesia," databoks. Accessed: Jan. 29, 2024. [Online]. Available: <https://databoks.katadata.co.id/datapublish/2019/10/08/persebaran-prevalensi-skizofreniapsikosis-di-indonesia>
- [2] S. N. Tarmizi, "Kemenkes Perkuat Jaringan Layanan Kesehatan Jiwa di Seluruh Fasyankes – Sehat Negeriku," Redaksi Sehat Negeriku. Accessed: Mar. 02, 2023. [Online]. Available: <https://sehatnegeriku.kemkes.go.id/baca/umum/20221010/4041246/kemenkes-kembangkan-jejaring-pelayanan-kesehatan-jiwa-di-seluruh-fasyankes/>

- [3] C. M. Annur, “Pengguna Twitter di Indonesia Capai 24 Juta hingga Awal 2023, Peringkat Berapa di Dunia?,” Katadata Media Networks. Accessed: Mar. 08, 2023. [Online]. Available: <https://databoks.katadata.co.id/datapublish/2023/02/27/pengguna-twitter-di-indonesia-capai-24-juta-hingga-awal-2023-peringkat-berapa-di-dunia>
- [4] M. Habibi and K. Kusumaningtyas, “Customer Experience Analysis Skincare Products Through Social Media Data Using Topic Modeling and Sentiment Analysis,” *JOURNAL OF SCIENCE AND APPLIED ENGINEERING*, vol. 6, no. 1, pp. 1–9, Jun. 2023, doi: 10.31328/JSAE.V6I1.4169.
- [5] A. G. Reece, A. J. Reagan, K. L. M. Lix, P. S. Dodds, C. M. Danforth, and E. J. Langer, “Forecasting the onset and course of mental illness with Twitter data,” *Scientific Reports 2017 7:1*, vol. 7, no. 1, pp. 1–11, Oct. 2017, doi: 10.1038/s41598-017-12961-9.
- [6] K. Chanda, S. Roy, H. Mondal, and R. Bose, “To Judge Depression and Mental Illness on Social Media Using Twitter,” *Univers J Public Health*, vol. 10, no. 1, pp. 116–129, Feb. 2022, doi: 10.13189/UJPH.2022.100113.
- [7] S. C. Guntuku, D. B. Yaden, M. L. Kern, L. H. Ungar, and J. C. Eichstaedt, “Detecting depression and mental illness on social media: an integrative review,” *Curr Opin Behav Sci*, vol. 18, pp. 43–49, Dec. 2017, doi: 10.1016/J.COBEHA.2017.07.005.
- [8] M. Habibi and P. W. Cahyo, “Clustering User Characteristics Based on the influence of Hashtags on the Instagram Platform,” *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, vol. 13, no. 4, pp. 399–408, 2019, doi: 10.22146/ijccs.50574.
- [9] P. W. Cahyo, “Klasterisasi Tipe Pembelajar Sebagai Parameter Evaluasi Kualitas Pendidikan di Perguruan Tinggi,” *Teknomatika*, vol. 11, no. 1, pp. 49–55, 2018.
- [10] F. B. P. Kencana, R. A. Yuana, and N. A. Pambudi, “Development of a Decision Support System for Clustering Scientific Publications Using K-Means,” *International Journal of Progressive Sciences and Technologies*, vol. 27, no. 2, pp. 773–783, Jul. 2021, doi: 10.52155/ijpsat.v27.2.3322.
- [11] D. Abdullah, S. Susilo, A. S. Ahmar, R. Rusli, and R. Hidayat, “The application of K-means clustering for province clustering in Indonesia of the risk of the COVID-19 pandemic based on COVID-19 data,” *Qual Quant*, vol. 56, no. 3, pp. 1283–1291, Jun. 2022, doi: 10.1007/S11135-021-01176-W/FIGURES/3.
- [12] M. Habibi and Sumarsono, “Implementation of Cosine Similarity in an automatic classifier for comments,” *JISKA (Jurnal Informatika Sunan Kalijaga)*, vol. 3, no. 2, pp. 38–46, 2018.
- [13] A. F. Hidayatullah and M. R. Maarif, “Penerapan Text Mining dalam Klasifikasi Judul Skripsi,” in *Seminar Nasional Aplikasi Teknologi Informasi (SNATi) Agustus*, Yogyakarta, 2016, pp. 1907–5022.
- [14] M. Habibi and P. W. Cahyo, “Journal Classification Based on Abstract Using Cosine Similarity and Support Vector Machine,” 2020. doi: <http://dx.doi.org/10.14421/jiska.2020.%25x>.

-
- [15] P. W. Cahyo and M. Habibi, "Clustering followers of influencers accounts based on likes and comments on Instagram Platform," *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, vol. 14, no. 2, pp. 199–208, 2020, doi: 10.22146/ijccs.53028.
- [16] U. A. Nasron and M. Habibi, "Analysis of Marketplace Conversation Trends on Twitter Platform Using K-Means," *Compiler*, vol. 9, no. 1, pp. 51–61, 2020, doi: 10.28989/compiler.v9i1.579.
- [17] H. Liu, I. Chatterjee, M. Zhou, X. S. Lu, and A. Abusorrah, "Aspect-Based Sentiment Analysis: A Survey of Deep Learning Methods," *IEEE Trans Comput Soc Syst*, vol. 7, no. 6, pp. 1358–1375, Dec. 2020, doi: 10.1109/TCSS.2020.3033302.
- [18] R. Diouf, E. N. Sarr, O. Sall, B. Birregah, M. Bousso, and S. N. Mbaye, "Web Scraping: State-of-the-Art and Areas of Application," *Proceedings - 2019 IEEE International Conference on Big Data, Big Data 2019*, pp. 6040–6042, Dec. 2019, doi: 10.1109/BIGDATA47090.2019.9005594.
- [19] M. Habibi, M. R. Ma'arif, and D. Subekti, "The Development of Social Media Intelligence System for Citizen Opinion and Perception Analysis over Government Policy," *Telematika : Jurnal Informatika dan Teknologi Informasi*, vol. 19, no. 1, pp. 31–46, Jul. 2022, doi: 10.31315/TELEMATIKA.V19I1.6447.
- [20] Y. Lv, "Student Behavior Analysis System Based on Kmeans Algorithm Under the Background of Smart Campus," *Proceedings - 2022 International Conference on Computer Network, Electronic and Automation, ICCNEA 2022*, pp. 174–177, 2022, doi: 10.1109/ICCNEA57056.2022.00047.
- [21] R. Y. Rajesh and G. Sindhu, "Predict the Game Analysis of Cricket Match Winning Using K-Nearest Neighbor and Compare Prediction Accuracy Over Support Vector Machine," *Proceedings - 2022 4th International Conference on Advances in Computing, Communication Control and Networking, ICAC3N 2022*, pp. 685–689, 2022, doi: 10.1109/ICAC3N56670.2022.10074193.
- [22] P. Ebrahimi, A. Salamzadeh, M. Soleimani, S. M. Khansari, H. Zarea, and M. Fekete-Farkas, "Startups and Consumer Purchase Behavior: Application of Support Vector Machine Algorithm," *Big Data and Cognitive Computing 2022, Vol. 6, Page 34*, vol. 6, no. 2, p. 34, Mar. 2022, doi: 10.3390/BDCC6020034.
- [23] M. Bansal, A. Goyal, and A. Choudhary, "A comparative analysis of K-Nearest Neighbor, Genetic, Support Vector Machine, Decision Tree, and Long Short Term Memory algorithms in machine learning," *Decision Analytics Journal*, vol. 3, p. 100071, Jun. 2022, doi: 10.1016/J.DAJOUR.2022.100071.
- [24] P. Nagaraj *et al.*, "Automatic and Adaptive Segmentation of Customer in R framework using K-means Clustering Technique," *2022 International Conference on Computer Communication and Informatics, ICCCI 2022*, 2022, doi: 10.1109/ICCCI54379.2022.9741067.
- [25] S. Wulandari, "Pentingnya Kesadaran tentang Kesehatan Mental - Diskominfo Prov. Kaltim," *Diskominfo Kaltim*. Accessed: Jun. 01, 2023. [Online]. Available:
-

<https://diskominfo.kaltimprov.go.id/index.php/kesehatan/pentingnya-kesadaran-tentang-kesehatan-mental>

- [26] Di. Nita, “4 Langkah ke Psikolog atau Psikiater Pakai BPJS Kesehatan, Akses Pengobatan Gratis,” *kompas.com*. Accessed: Jun. 01, 2023. [Online]. Available: <https://www.kompas.tv/article/387170/4-langkah-ke-psikolog-atau-psikiater-pakai-bpjs-kesehatan-akses-pengobatan-gratis>