Pemetaan Topik Pembicaraan Pada Komentar *Live* Youtube Menggunakan *K-Means Clustering* sebagai Identifikasi awal Kejahatan Verbal *Cyberbullying*

Alfirna Rizqi Lahitani[1*], Adlia Nur Zhafarina[2], Nanda Saputri Windi Oktavia[3], Nita Jariyah[4]

[1,3,4] Teknologi Informasi,Fakultas Teknik dan Teknologi Informasi Universitas Jenderal Achmad Yani Yogyakarta
[2] Hukum, Fakultas Ekonomi dan Sosial Universitas Jenderal Achmad Yani Yogyakarta
Jln. Siliwangi Ringroad Barat, Yogyakarta 55293 INDONESIA
Email: [1*]alfirnarizqi@gmail.com, [2]adliazhafarina@gmail.com, [3]sapnanda9@gmail.com, [4]neethajhary@gmail.com

Abstract*— The APJII 2022 survey noted that 98.02% of the internet is used by generation Z young people to access social media. The existence of social media certainly has a social impact, one of its as a place of bullying. Problems arise in the process of identifying victims and perpetrators of cyberbullying, besides difficulties in the meaning of words that contain elements of bullying can cause multiple interpretations that affect the investigation process. To overcome these problems, an initial identification effort is needed by grouping topics of conversation using the text mining method. This method can be started by preprocessing techniques, because the conversation data is in the text form. Next, topic grouping is carried out using the K-means clustering algorithm. Based on the results of the calculation, comment groups containing bullying are then mapped with Rapidminer tool. The contribution of this research is as the first step in the implementation of cybersecurity at the stage of identifying digital evidence for investigation purposes. This research resulted in mapping in the form of clusters. The results of the analysis on 2 clusters showed a grouping pattern in certain terms that were consistently grouped and contained the same term, namely the term "Nanyi", so that it can be concluded early that the topic being discussed in the Video is about singing activities or elements of conversation in the context of singing. While in other clusters, the terms in groups are more diverse because of the uniqueness of each term.

***Intisari*— Survey APJII 2022 mencatat 98,02% internet digunakan oleh kalangan anak muda generasi Z untuk mengakses media sosial. Keberadaan media sosial tentu saja memiliki dampak sosial, salah satunya sebagai tempat perundungan atau *bullying*. Permasalahan muncul dalam proses identifikasi korban dan pelaku *cyberbullying*, selain itu kesulitan dalam makna dan maksud kata yang mengandung unsur *bullying* dapat menimbulkan multitafsir yang mempengaruhi dalam proses investigasi. Untuk mengatasi permasalahan tersebut, diperlukan sebuah upaya identifikasi awal dengan mengelompokkan topik pembicaraan menggunakan metode *text mining*. Cara ini dapat dimulai dengan melakukan teknik *preprocessing*, dikarenakan data percakapan berbentuk teks. Selanjutnya dilakukan pengelompokkan topik menggunakan algoritma *K-means clustering*. Berdasarkan hasil perhitungan kelompok komentar yang mengandung perundungan kemudian dipetakan dengan bantuan *tools* Rapidminer. Kontribusi dari penelitian ini adalah sebagai langkah awal implementasi keamanan siber pada tahap identifikasi bukti digital untuk keperluan investigasi dan persidangan. Penelitian ini menghasilkan pemetaan dalam bentuk klaster. Hasil analisis pada 2 klaster terlihat pola pengelompokkan pada *term* tertentu yang secara konsisten terkelompok dan berisi *term* yang sama yaitu *term* "Nanyi", sehingga dapat ditarik kesimpulan awal bahwa topik yang sedang menjadi pembahasan dalam Video adalah seputar kegiatan bernyanyi atau unsur perbincangan dalam konteks bernyanyi. Sedangkan pada klaster lain term yang berkelompok lebih beragam karena keunikan masing-masing *term*.**

***Kata Kunci*— *Cyberbullying*, *Clustering*, K-Means, *Topic Modelling*, Youtube**

I. PENDAHULUAN

Internet menjadi insfrastrktur utama menuju dunia siber. Survey APJII 2022 mencatat 98,02% internet digunakan untuk mengakses media sosial dan penggunanya didominasi oleh kalangan generasi Z, dimana media yang paling sering dikunjungi adalah Youtube sebesar 63,02%[1]. Media sosial adalah bentuk kemajuan teknologi yang mempengaruhi masyarakat dalam aspek gaya hidup, cara pandang dan budaya[2]. Di kalangan anak muda generasi Z media sosial yang popular seperti Tiktok, Instagram, YouTube.

Media sosial yang paling banyak dikunjungi salah satunya adalah YouTube[1]. Layanan yang diberikan YouTube sangat beragam dengan interaksi komentar dan *emoticon* sehingga pengguna dapat mengekspresikan dirinya secara interaktif. Fitur komentar dapat ditemukan di video, YouTube Live, dan YouTube Shorts. Keberadaan media sosial tentu saja memiliki dampak positif dan negatif, sebagai salah satu dampak negatif dari keberadaan media sosial yaitu menggunakannya sebagai tempat perundungan atau *bullying*.

Bullying merupakan perilaku psikologis yang menjadi kejahatan verbal dan lebih menyakitkan dibandingkan kekerasan fisik seperti melecehkan, menghina, mengancam, merendahkan, atau menyakiti dan berujung pada penghinaan terhadap orang lain[3]. Perundungan yang dilakukan dalam aktivitas siber seperti komentar negatif pada postingan, pesan personal yang tak baik, serta mengolok-olok yang aktivitas tersebut terjadi di dunia siber disebut sebagai *cyberbullying*.

Merujuk pada undang-undang nomor 11 Tahun 2008 tentang Informasi dan Transaksi Elektronik (UU ITE). Tindakan menunjukan penghinaan terhadap orang lain tercermin pada pasal 27 ayat (3) UU ITE yang berbunyi: "Setiap orang dengan sengaja dan tanpa hak mendistribusikan dan/atau mentransmisikan dan/atau membuat dapat diaksesnya informasi elektronik dan/atau dokumen elektronik yang memiliki muatan penghinaan dan/atau pencemaran nama baik".

Permasalahan muncul dalam proses identifikasi korban dan pelaku *cyberbullying*, selain itu kesulitan dalam makna dan maksud kata yang mengandung unsur *bullying* dapat menimbulkan multitafsir yang mempengaruhi dalam proses investigasi. Untuk mengatasi permasalahan tersebut, dibutuhkan keahlian di bidang forensika digital, tujuannya untuk mendukung langkah investigasi dan pencarian barang bukti pada kejahatan *cyberbullying*.

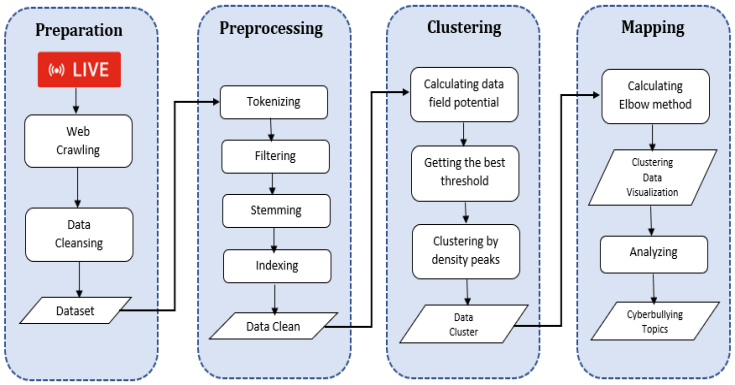
Sejumlah penelitian yang berkaitan dengan media soaial dan *cybersecurity*, dilakukan pada data Twitter[4] dan Youtube[5]. Maraknya kasus kejahatan melalui *platform* media sosial ditindaklanjuti secara khusus menginvestigasi kelompok kata sebagai alat bukti tindak kejahatan verbal[6]. Tindak kejahatan verbal *bullying* dapat mengancam pelakunya sebagai pertanggungjawaban pidana[7]. Sebagian besar analisis kata yang mengandung *cyberbullying* dalam bentuk teks, sehingga diolah menggunakan *textmining*, seperti penelitian[8] yang mengaplikasikan algoritma *K-means clustering* menggunakan *tools* Rapidminer. Penelitian dengan *textmining* untuk menganalisis pengaruh sebuah hastag di media sosial[9]. *Textminig* juga digunakan untuk pengelompokkan topik menggunakan algoritma *clustering*[10], pengelompokkan data dapat menjadi kontribusi dalam mengetahui topik[11]. Algoritma *clustering* dapat dikombinasikan dengan *cosine similarity*[12], maupun TF-IDF[13].

K-means merupakan metode yang bekerja dengan cara membagi himpunan data ke dalam kelompok, penentuan kelompok dilihat dari jarak terdekat antara data dengan *centroid*. Penelitian dengan algotirma K-Means *clustering* sebagai cara identifikasi awal *cyberbullying*[14],[3],[15],[16],[17]. Untuk mengukur kualitas dan optimalisasi *cluster* digunakan Elbow method[18],[19],[20],[21] . Berdasarkan sejumlah studi literatur yang telah dibahas, media sosial merupakan objek yang populer untuk diekplorasi, khususnya komentar di Youtube. Selain memiliki fitur komentar secara *asyncrhronous*, YouTube juga dilengkapi dengan fitur komentar secara langsung (live). Tidak jarang pada komentar ditemukan percakapan yang tidak sesuai bahkan menggunakan kata-kata yang kurang pantas.

Pada penelitian ini, agar dapat mendeteksi kejahatan *cyberbullying* dalam sebuah percakapan pada komentar di live YouTube, diperlukan sebuah upaya identifikasi awal dengan mengelompokkan topik pembicaraan menggunakan metode *text mining*. Dikarenakan data percakapan berbentuk teks, cara ini dapat dimulai dengan melakukan teknik *preprocessing*. Selanjutnya dilakukan pengelompokkan topik berdasarkan kata yang sudah teridentfikasi menggunakan algoritma K-means *clustering*. Berdasarkan hasil perhitungan kelompok komentar yang mengandung perundungan kemudian dipetakan atau dimodelkan dengan bantuan *tools* Rapidminer. Kontribusi dari penelitian ini adalah sebagai langkah awal implementasi keamanan siber pada tahap identifikasi bukti digital untuk keperluan investigasi dan persidangan.

II. METODE

Penelitian ini menggunakan data berbasis teks yang dihimpun berasal dari komentar live pada sebuah kanal Youtube. Adapun tahap dalam penelitian ini seperti yang ditunjukkan pada Gambar 1.



Gambar. 1 Tahap Penelitian

Berdasarkan Gambar 1, tahapan dalam penelitian ini meliputi 4 tahap sebagai berikut:

A. *Tahap Persiapan*

Pada tahap ini telah dilakukan studi literatur untuk kajian awal penelitian. Selanjutnya akan dilakukan penentuan sumber data berupa channel Youtube untuk dilakukan tahap pengumpulan Data. Data yang dihimpun merupakan data komentar berbasis teks. Proses menghimpun data menggunakan teknik *crawling*, yaitu proses pemindaian sebuah konten pada *website*. Data yang telah dapatkan akan filter terlebih dahulu dari *noise* untuk siap menjadi dataset dalam penelitian.

B. *Tahap Pengolahan Data (Preprocessing)*

Data dalam bentuk teks yang telah dihimpun akan melalui beberapa proses, tujuan proses ini adalah ekstraksi data dengan *preprocessing* yang meliputi *Tokenizing, Filtering, Stemming* dan *Indexing*. Hasil akhir *preprocessing* berupa data bersih yang siap diolah[22].

C. *Tahap Analisis Data (Clustering)*

Data yang telah diekstraksi kemudian dianalisis dengan metode *clustering* menggunakan algoritma *K-means clustering* sebagai formula dalam menghitung kelompok kata. *Clustering* merupakan teknik *data mining* yang mengelompokkan data berdasarkan pada objek/karakteristik yang sama kemudian dikumpulkan dalam satu kluster dan klaster tersebut berbeda dengan klaster lainnya. *K-Means* digunakan untuk mengelompokkan karakter berdasarkan karakter yang telah didefinisikan. Persamaan *K-Means* seperti yang ditunjukkan pada persamaan 1:[9]

$$d(X_j, C_j) = \sqrt{\sum_{j=1}^{n}(X_j - C_j)^2} \qquad (1)$$

Keterangan:
d = distance
n = number of objects
j = start from 1 to n
Xj = feature object to j with respect to x
Cj = centroid feature to j

D. *Tahap Pemetaan Hasil (Mapping)*

Hasil kluterisasi akan dipetakan dengan bantuan *tools* Rapidminer dalam bentuk visual *scatter* diagram maupun *wordcloud*.

III. HASIL DAN PEMBAHASAN

Hasil penelitian ini berupa pemetaan kelompok kata sebagai identifikasi awal *cyberbullying*. Adapun hasil dari masing-masing tahapan sebagai berikut:

A. *Tahap Persiapan*

Pada penelitian ini, pengumpulan data berasal dari komentar pada sebuah YouTube yang merupakan data berbasis teks. Pengumpulan data dilakukan dengan teknik *crawling* yang bertujuan mengambil data tertentu, adapun data komentar berasalh dari channel berikut: https://www.youtube.com/shorts/RABxiBvgxpU. Hasil *crawling* diperoleh 100 data yang berisi komentar yang akan menjadi sumber data untuk diolah dalam ranah *text mining*.

B. *Tahap Pengolahan Data (Preprocessing)*

Hasil *crawling* disimpan dalam bentuk .csv untuk selanutnya dioleh dengan teknik *preprocessing* yang terdiri dari: *case folding*, *tokenizing*, *stopword removal*, *stemming* dan *indexing*. Proses *preprocessing* menggunakan bahasa Phyton. Berikut tampilan *preprocessing* yang ditunjukkan pada Gambar 2.

```
[ ] import pandas as pd
    import numpy as np

    tweet_data = pd.read_csv("data.csv")

    tweet_data.head()
```

	tweet	Unnamed: 1
0	Ini sih masalah teknik. Berarti tehnik bernyan...	view comment
1	Sebenarnya suara nih orang tuh bangus banget.....	view comment
2	Dasar yang edit dan buat cerita pinter banget!...	view comment
3	Jam terbang GK pernah bohong ..	view comment
4	Nada tinggi itu susah	view comment

```
tweet_data['tweet'] = tweet_data['tweet'].str.lower()

print('Case Folding result : \n')
print(tweet_data['tweet'].head(5))
print('\n\n\n')
```

Gambar. 2 Tahap *Preprocessing*

C. *Tahap Analisis Data (Clustering)*

Data yang telah di*preprocessing* kemudian disimpan dalam format .csv. Diklaster menggunakan *tools* Rapidminer dengan pengujian 2 kluster, 3 kluster, 5 kluster, 7 kluster dan 10 kluster. Berikut adalah cuplikan masing-masing perhitugan pada Tabel 1:

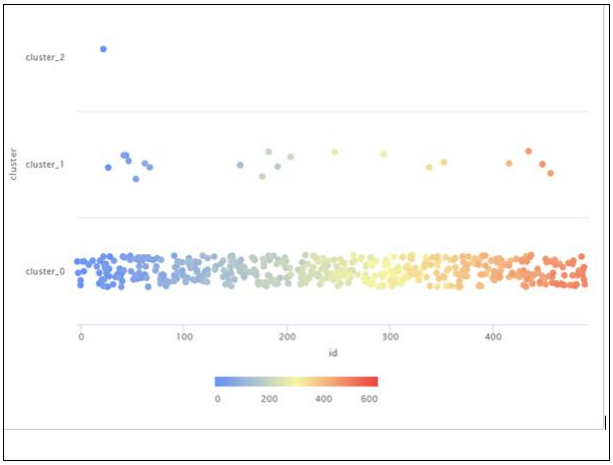
Tabel. 1 Hasil Klaster

No	Kluster	Total Item	Keterangan
1	2	488	Cluster 0: 468 items Cluster 1: 20 items
2	3	488	Cluster 0: 467 items Cluster 1: 20 items Cluster 2: 1 items
3	5	488	Cluster 0: 464 items Cluster 1: 20 items Cluster 2: 1 items Cluster 3: 1 items Cluster 4: 2 items
4	7	488	Cluster 0: 456 items Cluster 1: 20 items Cluster 2: 3 items Cluster 3: 1 items Cluster 4: 6 items Cluster 5: 1 items Cluster 6: 1 items
5	10	488	Cluster 0: 422 items Cluster 1: 15 items Cluster 2: 16 items Cluster 3: 20 items Cluster 4: 6 items Cluster 5: 1 items Cluster 6: 2 items Cluster 7: 5 items Cluster 8: 4 items Cluster 9: 2 items Cluster 10: 1 items

D. *Tahap Pemetaan Hasil (Mapping)*

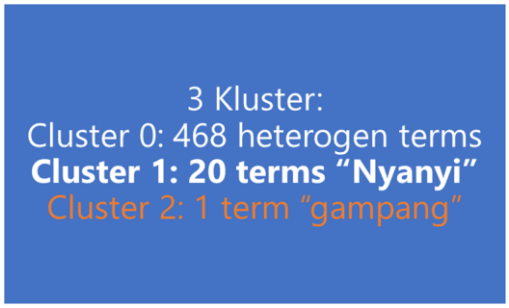
Hasil klastering yang dilakukan bertujuan untuk menemukan kelompok kata yang memiliki kesamaan makna atau konteks dikelompokkan dalam klaster yang sama. Setiap kelompok kata yang terbentuk oleh algoritma akan diwakili oleh pusat klaster (*centroid*) dan berisi sejumlah kata. kata-kata yang paling dekat dengan pusat klaster adalah yang paling mendefinisikan klaster tersebut. Menemukan pola dan hubungan antara kata-kata dalam teks. Kesamaan makna kata dianalisis dari *term* yang paling sering muncul bersama dalam teks, sehingga apabila *term* tersebut memiliki makna yang sama, maka akan terkelompok dalam kluster yang sama. Dalam konteks ini dapat diidentifikasi tema atau topik tertentu dan mengidentifikasi pola sentimen.

Berikut adalah hasil pemetaan *term* yang divisualisasikan dalam bentuk diagram *scatter* menggunakan Rapidminer yang ditunjukkan pada Gambar 3.



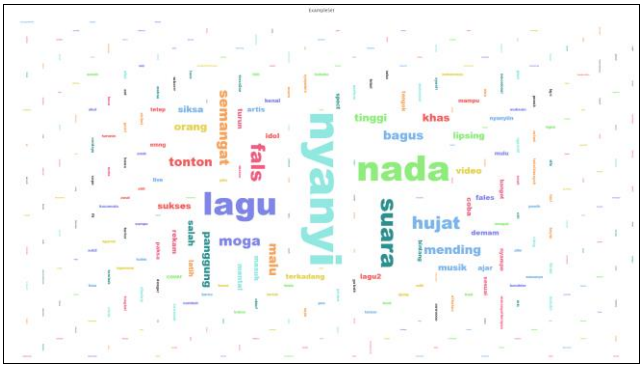
Gambar. 3 Visualisasi Klaster dengan Scatter

Hasil klaster yang ditunjukkan pada Gambar 3 menerangkan bahwa pada klaster 0 terdapat lebih dari 400 items yang terkelompok menjadi satu grup. Pada klaster 1 terdapat lebih sedikit item yang terkelompok yang memiliki kemiripan. Sedangkan pada klaster 2 hanya terdapat satu item. Adapun data term pada masing-masing kelompok seperti yang ditunjukkan pada Gambar 4.



Gambar. 4 Data item pada percobaan 3 klaster

Berdasarkan klasterisasi percobaan 3 klaster, ingin diketahui kelompok kata yang terkandung dalam komentar video, divisualisasikan menjadi *wordcloud* seperti yang ditunjukkan pada Gambar 5.



Gambar. 5 Visualisasi Wordcloud.

Berdasarkan hasil visualisasi pada Gambar 4 dan Gambar 5, pengelompokkan term dibagi menjadi 3 klaster dan diperoleh Cluster 0: 467 items, Cluster 1: 20 items, Cluster 2: 1 item. Pada klaster 1 dan 2 terlihat pengelompokkan pada *term* tertentu seperti pada klater 1 yang berisi kumpulan *term* yang sama yaitu term "Nanyi", sehingga dapat disimpulkan bahwa topik yang sedang menjadi pembahasan dalam Video adalah seputar kegiatan bernyanyi atau unsur-unsur yang lekat dengan perbincangan dalam konteks bernyanyi. Sedangkan pada klaster 0, *term* yang berkelompok lebih beragam dikarenakan setiap *term* unik sehingga pada seluruh *term* dengan keunikan masing-masing dikelompokkan menjadi klaster 0. Makna dari kata-kata yang terkandung dalam klaster 0 bervariasi mulai dari term yang mengandung unsur sentimen negatif namun juga terdapat unsur sentimen yang positif, sehingga untuk hal ini perlu dilakukan pengamatan lebih lanjut.

IV. KESIMPULAN

Analisis Pada klaster terlihat pola pengelompokkan pada *term* tertentu yang secara konsisten terkelompok dan muncul berisi kumpulan term yang sama yaitu term "Nanyi", sehingga dapat ditarik kesimpulan awal bahwa topik yang sedang menjadi pembahasan dalam Video adalah seputar kegiatan bernyanyi atau unsur-unsur yang lekat dengan perbincangan dalam konteks bernyanyi. Sedangkan pada kalaster lainnya, *term* yang berkelompok lebih beragam. Makna dari kata-kata yang terkandung dalam klaster yang memiliki *term* yang lebih bervariasi mengandung unsur negatif namun juga terdapat unsur yang positif, dengan demikian pada klaster inilah dapat dijadikan sebagai dasar dilakukan pengamatan lebih lanjut untuk inversigasi.

UCAPAN TERIMA KASIH

Ucapan terima kasih ini ditujukan kepada Direktorat Riset, Teknologi, dan Pengabdian kepada Masyarakat (DRTPM) Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi (Kemdikbudristek) Republik Indonesia yang telah memberikan kesempatan kepada tim peneliti untuk melakukan penelitian pada tahun 2023 dengan skema PDP.

REFERENSI

[1] APJII, "Profil Internet Indonesia 2022," *Apji.or.Od*, no. June, p. 10, 2022, [Online]. Available: apji.or.id.

[2] L. Fazry and N. C. Apsari, "Pengaruh Media Sosial Terhadap Perilaku Cyberbullying Di Kalangan Remaja," *J. Penelit. dan Pengabdi. Kpd. Masy.*, vol. 2, no. 2, p. 272, 2021, doi: 10.24198/jppm.v2i2.34679.

[3] A. Muhariya, "Pengelompokkan Komentar Pada Media Sosial Instagram Menggunakan Metode K-Means Clustering Untuk Identifikasi Awal Cyberbullying," pp. 1–71, 2022.

[4] L. Siagian, A. Budiarto, P. Strategi, P. Udara, and U. Pertahanan, "PERAN KEAMANAN SIBER DALAM MENGATASI KONTEN NEGATIF GUNA MEWUJUDKAN KETAHANAN INFORMASI NASIONAL," pp. 1–18, 2017.

[5] A. R. Lahitani, U. S. Aesyi, N. Wulandari, and B. D. Santosa, "Cosine Similarity untuk Mengukur Tingkat Kesadaran pada Topik Software Security Berbasis Teks Komentar di Media Sosial Youtube," vol. 8, no. November, pp. 107–115, 2022, doi: 10.34128/jsi.v8i2.535.

[6] E. Nurdiyanto and G. A. Resticka, "Kata Makian Sebagai Alat Bukti Investigasi Tindak Kejahatan Verbal (Kajian Linguistik Forensik Berbasis Semantik dan Pragmatik)," *Gramatika STKIP PGRI Sumatera Barat*, vol. 7, no. 1, pp. 43–56, 2021.

[7] A. Linda Mustika, M. Santoso, N. Sabrina, and K. Penulis, "Pertanggungjawaban Pidana Atas Tindak Pidana Pelecehan Verbal Melalui Media Sosial," *Bhirawa Law J.*, vol. 2, no. 1, 2021, [Online]. Available: http://jurnal.unmer.ac.id/index.php/blj/.

[8] B. Rahmat *et al.*, "Implemetasi k-means clustering pada rapidminer untuk analisis daerah rawan kecelakaan," *Semin. Nas. Ris. Kuantitatif Terap. 2017*, no. April, pp. 58–60, 2017, [Online]. Available: https://ojs.innov-

center.org/index.php/snrkt2017/article/download/10/9.
[9] M. Habibi and P. W. Cahyo, "Clustering User Characteristics Based on the influence of Hashtags on the Instagram Platform," *IJCCS (Indonesian J. Comput. Cybern. Syst.*, vol. 13, no. 4, p. 399, 2019, doi: 10.22146/ijccs.50574.
[10] H. Irsyad and M. R. Pribadi, "Implementasi Text Mining Dalam Pengelompokan Data Tweet Pertanian Indonesia Dengan K-Means," *Kurawal - J. Teknol. Inf. dan Ind.*, vol. 3, no. 2, pp. 164–172, 2020, doi: 10.33479/kurawal.v3i2.347.
[11] F. Kolini and L. Janczewski, "Clustering and topic modelling: A new approach for analysis of national cybersecurity strategies," *Proc. ot 21st Pacific Asia Conf. Inf. Syst. "'Societal Transform. Through IS/IT'", PACIS 2017*, 2017.
[12] N. W. Utami and I. G. J. Eka Putra, "Text Minig Clustering Untuk Pengelompokan Topik Dokumen Penelitian Menggunakan Algoritma K-Means Dengan Cosine Similarity," *J. Inform. Teknol. dan Sains*, vol. 4, no. 3, pp. 255–259, 2022, doi: 10.51401/jinteks.v4i3.1907.
[13] R. Khan, Y. Qian, and S. Naeem, "Extractive based Text Summarization Using KMeans and TF-IDF," *Int. J. Inf. Eng. Electron. Bus.*, vol. 11, no. 3, pp. 33–44, 2019, doi: 10.5815/ijieeb.2019.03.05.
[14] K. P. Sinaga and M. S. Yang, "Unsupervised K-means clustering algorithm," *IEEE Access*, vol. 8, pp. 80716–80727, 2020, doi: 10.1109/ACCESS.2020.2988796.
[15] Aderibigbe, "Pemodelan Deteksi Cyber Bullying Pada Jejaring Sosial Twitter," *Energies*, vol. 6, no. 1, pp. 1–8, 2018, [Online]. Available: http://journals.sagepub.com/doi/10.1177/1120700020921110%0Ahttps://doi.org/10.1016/j.reuma.2018.06.001%0Ahttps://doi.org/10.1016/j.arth.2018.03.044%0Ahttps://reader.elsevier.com/reader/sd/pii/S1063458420300078?token=C039B8B13922A2079230DC9AF11A333E295FCD8.
[16] M. Ahmed, R. Seraj, and S. M. S. Islam, "The k-means algorithm: A comprehensive survey and performance evaluation," *Electron.*, vol. 9, no. 8, pp. 1–12, 2020, doi: 10.3390/electronics9081295.
[17] N. F. Hasan, "Deteksi Cyberbullying pada Facebook Menggunakan Algoritma K-Nearest Neighbor," *J. Smart Syst.*, vol. 1, no. 1, pp. 35–44, 2021, doi: 10.36728/jss.v1i1.1605.
[18] H. Humaira and R. Rasyidah, "Determining The Appropiate Cluster Number Using Elbow Method for K-Means Algorithm," 2020, doi: 10.4108/eai.24-1-2018.2292388.
[19] M. A. Syakur, B. K. Khotimah, E. M. S. Rochman, and B. D. Satoto, "Integration K-Means Clustering Method and Elbow Method for Identification of1. Syakur MA, Khotimah BK, Rochman EMS, Satoto BD. Integration K-Means Clustering Method and Elbow Method for Identification of the Best Customer Profile Cluster. IOP Conf Ser Mat," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 336, no. 1, 2018, doi: 10.1088/1757-899X/336/1/012017.
[20] M. Cui, "on the Elbow Method," pp. 5–8, 2020, doi: 10.23977/accaf.2020.010102.
[21] J. D'Silva and U. Sharma, "Unsupervised Automatic Text Summarization of Konkani Texts using K-means with Elbow Method," *Int. J. Eng. Res. Technol.*, vol. 13, no. 9, pp. 2380–2384, 2020, doi: 10.37624/ijert/13.9.2020.2380-2384.
[22] A. R. Lahitani, "Automated Essay Scoring menggunakan Cosine Similarity pada Penilaian Esai Multi Soal," *J. Kaji. Ilm.*, vol. 22, no. 2, pp. 107–118, 2022, doi: 10.31599/jki.v22i2.1121.