



Penerapan Metode K-Means Dalam Melakukan Pengelompokan Bencana Alam di Indonesia Dilakukan dengan Memanfaatkan Teknik Text Mining

Iqbal Alfian

Jurnal Algoritma
STMIK LIKMI

Jl. Ir. H. Juanda No.96, Lebakgede, Kecamatan Coblong, Kota Bandung, Jawa Barat 40132
Email : info@likmi.ac.id

if18.iqbalalfian@mhs.ubpkarawang.ac.id

Abstrak – Bencana alam di Indonesia beberapa diantaranya adalah banjir, gempa bumi, erupsi dan lainnya. Dalam menghadapi hal tersebut, pengelompokan jenis bencana menjadi sangat krusial untuk menetapkan langkah dan rencana yang sesuai. Teknologi dapat digunakan untuk memudahkan proses pengelompokan tersebut, salah satunya dengan memanfaatkan teknik *text mining*. Pengelompokan informasi dilakukan dengan memasukan beberapa kluster dengan dasar keterkaitan antar kata menggunakan algoritma *K-Means*. Dalam penelitian ini, bertujuan untuk menghasilkan sebuah model pengelompokan bencana alam di Indonesia dengan menerapkan algoritma *K-Means*. Analisa didasari oleh data komentar masyarakat tentang bencana alam di media sosial *Twitter*. Penggunaan metode *text mining* dengan aplikasi RStudio berhasil melakukan pengelompokan bencana alam berdasarkan potensi dan jenisnya dari data komentar masyarakat di media sosial *twitter*. Setelah melakukan text cleaning, text processing, dan metode TF-IDF, diketahui bahwa banjir dan gempa merupakan topik bencana alam tertinggi dari penambangan data tersebut. Metode *unsupervised* dengan algoritma K-Means digunakan untuk membangun kelompok topik berdasarkan jarak keterkaitan antar kata-kata. Evaluasi dilakukan menggunakan metode *Sum of Square Error* dan *Silhouette Coefficient*, dan diperoleh akurasi sebesar 75.0% dan 96.7%. Simpulan bahwa algoritma *K-Means* berhasil membangun kelompok topik berdasarkan jarak keterkaitan antar kata-kata pada data komentar masyarakat tentang bencana alam di *twitter*.

Kata Kunci – Bencana Alam; *Clustering*; *Data Mining*; *K-Means*; *Text Mining*; *Twitter*.

I. PENDAHULUAN

Salah satu teknologi saat ini yang dapat menjadi sumber data adalah media sosial seperti *Twitter*. Data tersebut dapat membantu peneliti untuk memperoleh gambaran yang lebih lengkap tentang bencana alam yang terjadi dan potensinya. Namun, jumlah data yang besar membuat analisis manual menjadi lebih lama. Perlunya metode pengolahan data yang dapat mempermudah analisis data besar tersebut. Metode yang dapat digunakan adalah teknik *text mining*, teknik tersebut memungkinkan melakukan analisis terhadap data teks dengan menggunakan algoritma tertentu untuk menghasilkan kelompok atau kategori tertentu.

Sebelumnya penelitian mengenai algoritma *K-Means* oleh [1], data yang digunakan diperoleh dari dataset mentah milik perusahaan dan metode yang digunakan adalah *K-Means* untuk mengelompokkan data barang berdasarkan tingkat penjualannya. Sementara itu, penelitian lainnya oleh [2] menggunakan metode *sentiment analysis* pada data *Twitter* dengan kata kunci "grab" dan "gojek" untuk mengelompokkan data berdasarkan tanggapan positif dan negatif dari masyarakat. Hasilnya menunjukkan akurasi sentiment objek Grab sebesar 74,34% dan Gojek sebesar 68,84%. Berikutnya penelitian tentang pengelompokan data transaksi [3], data didapatkan dari transaksi pembayaran CV Digital Dimensi. Metode *K-Means* dilakukan pada penelitian ini

sebagai pengelompok data yang telah didapatkan. Terakhir penelitian oleh [4], data melalui portal berita berupa teks dan diolah oleh metode K-Means serta mendapatkan nilai akurasi sebesar 72.01%.

Berdasarkan permasalahan tersebut, tujuan dari pembuatan penelitian ini ialah untuk mengidentifikasi dan mengklasterkan variasi bencana alam dengan memanfaatkan teknik text mining. Data diperoleh dari penambahan komentar masyarakat di media sosial *Twitter* yang berkaitan dengan bencana alam. Metode *text mining* digunakan untuk mengolah data tersebut dengan langkah-langkah *stemming*, *tokenizing*, *stopword removal*, dan *TF-IDF*. Melalui pengolahan data, didapatkan hasil bahwa banjir dan gempa merupakan topik bencana alam tertinggi dari data yang diperoleh. Pengelompokan data dilakukan menggunakan jenis *unsupervised* melalui Algoritma *K-Means* dan dievaluasi melalui metode *Silhouette Coefficient* dan *Sum of Square Error*. Penelitian ini dapat dijadikan sebagai pilihan rujukan untuk yang berkepentingan terkait dalam mengambil kebijakan dan tindakan dalam menghadapi bencana alam di Indonesia.

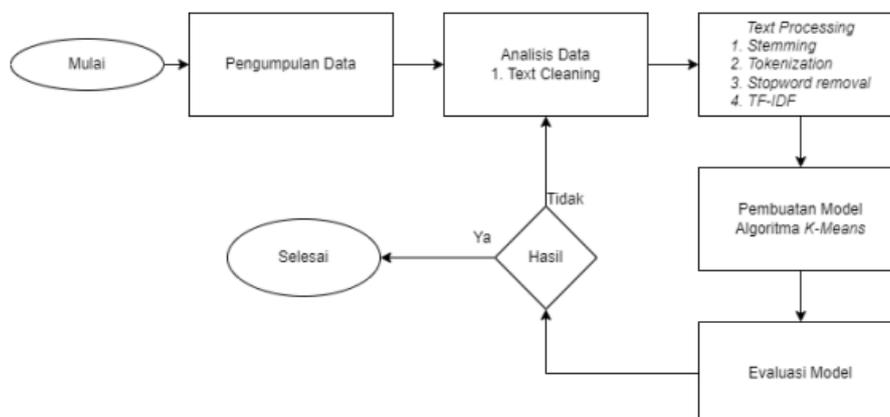
II. METODOLOGI PENELITIAN

A. Bencana Alam

Gunung berapi yang ada di Indonesia adalah salah satu yang terbanyak di seluruh dunia, sehingga akan membuat bencana jenis erupsi mempunyai jumlah terbanyak. Potensi bencana lainnya dapat disebabkan oleh cuaca serta iklim dan tingginya curah hujan, lalu potensi berikutnya yaitu kemarau panjang yang akan membuat bencana kekeringan [5]. Serangkaian peristiwa bencana merupakan salah satunya disebabkan oleh alam [6]. Namun, manusia juga berkontribusi pada meningkatnya frekuensi dan intensitas bencana alam, seperti melalui kerusakan lingkungan dan perubahan iklim yang menyebabkan cuaca ekstrem [7]. Bencana alam memiliki dampak yang signifikan terhadap kehidupan manusia, termasuk hilangnya nyawa, kerugian ekonomi, dan kerusakan lingkungan. Oleh karena itu, pengelompokan jenis bencana alam menjadi penting dalam upaya penanganan dan mitigasi risiko bencana alam. Penelitian ini dapat membantu Tim Penanggulangan bencana dalam memprediksi kejadian dimasa depan Bersama-sama dengan masyarakat. Dengan adanya pedoman dan prosedur pada tahap awal penerapan manajemen risiko akan memaksa individu pada berbagai tingkatan di organisasi untuk melaksanakan prosedur pengelolaan risiko organisasi [8].

B. Prosedur Penelitian

Kegiatan pertama yang dilakukan dalam penelitian ini adalah menganalisis dan mengumpulkan data, kedua melakukan text Processing, ketiga mengelompok data yang didapatkan lalu dilakukan evaluasi model [9]. Gambar 1 merupakan Flowchart prosedur penelitian.



Gambar 1: Prosedur Penelitian

C. Text Mining

Ada beberapa cara untuk mengidentifikasi pola dari suatu masalah, salah satunya menggunakan konsep *text mining*. Metode ini akan mengambil seluruh informasi yang berkaitan dengan kualitas terbaik mengacu pada pola serta keterkaitan antar kata untuk membuat pembelajaran pola-pola statistik [10]. Dalam penerapannya, *text mining* memproses data dalam bentuk kata-kata untuk mendapatkan informasi dan memiliki berbagai area penerapan yaitu ekstraksi informasi, lacak topik, perangkuman, kategorisasi topik, pengklasteran, memodelkan topik, dan penjawaban pertanyaan [11]. Proses yang dilakukan pada *text mining*, yaitu *Text Cleaning* dan *Text Processing* [12].

D. Algoritma K-Means

Algoritma pengelompokan yang menggunakan metode non-hierarki untuk membagi data ke dalam lebih dari satu kelompok yang memiliki keterkaitan [13]. Algoritma ini termasuk dalam kategori *unsupervised learning*, yang artinya ia mampu mempelajari pola-pola pada dataset tanpa memerlukan label [14].

Proses yang dilakukan *K-Means* adalah :

1. Buat kluster dengan jumlah tertentu
2. Untuk penghitungan jarak antara data dan centroid, gunakanlah formula *Euclidean distance*:

$$D_{ij} = \sqrt{(x_{1j} - v_{1j})^2 + \dots + (x_{ij} - v_{kj})^2} \quad (1)$$

Dimana : D_{ij} = jarak data i ke titik pusat *cluster* ke k , x_{ij} = data ke- i pada atribut ke- j , v_{kj} = titik pusat ke- k atribut ke- j .

3. Menghitung kembali pusat titik *cluster* (*centroid*) dengan anggota *cluster* yang baru, berikut rumus untuk menghitung *centroid*:

$$C_1 = \frac{1}{M} \sum_{j=1}^M x_j \quad (2)$$

4. Proses selesai jika pusat cluster tidak berubah, gunakan langkah ke 3 jika proses belum selesai.

E. Sum of Square Error

Metode *Sum of Square Error* (*SSE*) berfungsi untuk mengukur nilai akurasi dalam mendapatkan selisih dari total nilai sebenarnya $\text{Varian} = \text{SSE}/n$, proses menghitung varian untuk seluruh populasi, proses menghitung varian sample (3).

$$\begin{aligned} \text{Varian} &= \text{SEE}/((n-1)) & (3) \\ \text{Varian} &= 2,191/((7)) \\ \text{Varian} &= 0.313 \end{aligned}$$

F. Silhouette Coefficient

Evaluasi *Silhouette Coefficient* adalah himpunan dari dua metode yaitu metode *cohesion* dan *separation*. Metode *cohesion* untuk mengukur kedekatan keterkaitan antara objek data sebuah *cluster*. Lalu metode *separation* untuk mengukur antara jarak *cluster* terpisah dengan *cluster* lain. Kedua metode *Silhouette Coefficient* digunakan untuk validasi hasil *clustering*. Persamaan *Silhouette Coefficient* (4).

$$\text{sil}(c) = \text{sil}(k) \frac{1}{|k|} \sum_{i=1}^k \text{sil}(c_i) \quad (4)$$

III. HASIL DAN DISKUSI

A. Hasil Pengumpulan Data

Data diambil pada periode Januari hingga Maret 2023 dari media *Twitter* dengan total 4169 data yang memiliki dua atribut, yaitu nomor dan teks. Tabel 1 adalah hasil pengumpulan data.

Tabel 1: Hasil Pengumpulan Data *Twitter*

Nomor	Text
1.	RT @convomf: Teman-teman minta doanya ya untuk Almarhumah ibunda temanku. Mamanya temenku salah satu korban bencana longsor yg terjadi di Natuna pada hari senin lalu (06 maret 2023). Jenazahnya baru ditemukan tadi pagi, 4 hari setelah kejadian. Komunikasi terakhir temanku dengan (cont) https://t.co/5NTcYJOV9D
.	.
.	.
.	.
4169	Menurutku pesan Makoto Shinkai di film suzume tuh walau diberi bencana tetep percaya hari esok, kudu kuat, kudu punya harapan, tetep punya semangat hidup. Kalau yang tenki no ko ya walau di tengah bencana, Jepang tetep bisa jalan tetep punya harapan

B. Text Cleaning

Setelah data berhasil dikumpulkan, langkah selanjutnya adalah membersihkan teks dengan cara menghilangkan simbol, angka, dan URL. Tabel 2 menunjukkan hasil dari proses pembersihan teks (*text cleaning*).

Tabel 2: *Text Cleaning*

Nomor	Sebelum	Sesudah
1	RT @convomf: Teman-teman minta doanya ya untuk Almarhumah ibunda temanku. Mamanya temenku salah satu korban bencana longsor yg terjadi di Natuna pada hari senin lalu (06 maret komunikasi terakhir temanku dengan (cont) https://t.co/5NTcYJOV9D	RT @convomf: Teman-teman minta teman-teman minta doanya ya untuk almarhumah ibunda doanya ya untuk Almarhumah ibunda temanku. mamanya temenku salah satu korban bencana temanku. Mamanya temenku salah satu longsor yg terjadi di natuna pada hari senin lalu maret korban bencana longsor yg terjadi di jenazahnya baru ditemukan tadi pagi hari setelah kejadian Natuna pada hari senin lalu (06 maret komunikasi terakhir temanku dengan 2023). Jenazahnya baru ditemukan tadi pagi, 4 hari setelah kejadian. Komunikasi terakhir temanku dengan (cont) https://t.co/5NTcYJOV9D
-	-	-
-	-	-
-	-	-
4169	Menurutku pesan Makoto Shinkai di film suzume tuh walau diberi bencana tetep percaya hari esok, kudu kuat, kudu punya harapan, tetep punya semangat hidup. Kalau yang tenki no ko ya walau di tengah bencana, Jepang tetep bisa jalan tetep punya harapan	Menurutku pesan makoto shinkai di film suzume tuh walau suzume tuh walau diberi bencana tetep diberi bencana tetep percaya hari esok, kudu kuat, kudu percaya hari esok, kudu kuat, kudu punya punya harapan, tetep punya semangat hidup. kalau yang harapan, tetep punya semangat hidup. tenki no ko ya walau di tengah bencana, jepang tetep bisa Kalau yang tenki no ko ya walau di tengah jalan tetep punya harapan bencana, Jepang tetep bisa jalan tetep punya harapan

C. Text Processing

Langkah awal dalam *Text Processing* adalah *Stemming*, dimana kata-kata berimbuhan akan diganti dengan kata-kata dasar. Proses *Stemming* pada Tabel 3.

Tabel 3: *Stemming*

Nomor	Sebelum	Sesudah
1	teman-teman minta doanya ya untuk temanteman minta doa ya untuk almarhumah ibunda teman almarhumah ibunda temanku. mamanya mama temenku salah satu korban bencana longsor yg jadi temenku salah satu korban bencana di natuna pada hari senin lalu maret jenazah baru temu tadi longsor yg terjadi di natuna pada hari pagi hari setelah jadi komunikasi akhir teman dengan cont senin lalu maret jenazahnya baru ditemukan tadi pagi hari setelah kejadian komunikasi terakhir temanku dengan	
-	-	
-	-	
-	-	
4169	catatan materi gabungan ipa kimia fisika catat materi gabung ipa kimia fisika biolog kela kurikulum biolog kela kurikulum merdeka merdeka pemanasan global pemanasan global	

Tahap kedua dari *Text Processing* adalah *Tokenization*, dimana tujuannya adalah untuk memisahkan kata-kata dalam kalimat yang telah melalui tahap *Stemming*. Hasil dari proses *Tokenization* pada Tabel 4.

Tabel 4: Hasil *Tokenisasi*

Nomor	Sebelum	Sesudah
1	temanteman minta doa ya untuk almarhumah ibunda teman mama temenku salah satu korban bencana longsor yg jadi di natuna pada hari senin lalu maret jenazah baru temu tadi pagi hari setelah jadi komunikasi akhir teman dengan cont	Temanteman - - - cont
-	-	
-	-	
-	-	
4169	catat materi gabung ipa kimia fisika biolog kela kurikulum merdeka pemanasan global	catat - - - global

Proses ketiga dalam pengolahan teks adalah penghapusan *stopword* dengan tujuan menghilangkan kata-kata yang tidak berhubungan dengan objek yang sedang diteliti. Dalam proses ini, sebanyak 27.317 kata telah dihapus dan hasilnya akan ditunjukkan dalam Tabel 5. Nomor pada tabel menunjukkan urutan data, sedangkan kolom "sebelum" dan "setelah" menunjukkan data sebelum dan sesudah penghapusan *stopword* dilakukan.

Tabel 5: *Stopword Removal*

Nomor	Sebelum	Sesudah
1	pokok	pokok
2	yaaa	konflik
-	-	-
-	-	-
-	-	-
71140	berangkat	berangkat
71141	ke	negara

Tabel 6 adalah hasil perhitungan *Document Frequency* setelah melakukan proses keempat, yaitu perhitungan *Term Frequency*, *Inverse Document Frequency*, dan *Term Frequency/Inverse Document Frequency*.

Tabel 6: Frekuensi Kata

No.	words	frequency
1	banjir	1800
2	bencana	1228
3	gempa	1037
4	beliung	551
-	-	550
-	-	397
-	-	392
8	segment	1

Proses kelima melibatkan pembobotan kata-kata menggunakan metode *TF-IDF* [15], yang akan disajikan pada Tabel 7: Selanjutnya, bertujuan untuk mencari frekuensi keterkaitan pada Tabel 8.

Tabel 7: Hasil *TF-IDF*

word	n	total	tf	idf	tf_idf
banjir	1780	27810	0.064006	0	0
bencana	1306	23882	0.054686	0	0
gempa	1123	18099	0.062048	0	0
beliung	618	9485	0.065156	0.223144	0.014539
-	387	9485	0.040801	0.223144	0.009105
-	256	23882	0.010719	0	0
-	216	27810	0.007767	0	0
lahat	213	27810	0.007659	0.916291	0.007018

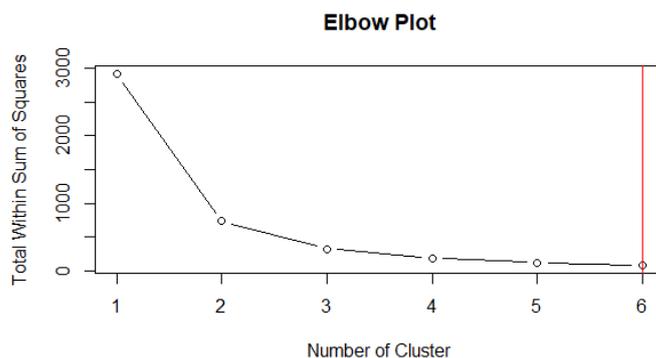
Tabel 8: Frekuensi Keterkaitan Kata

no	word	n
1	pute beliung	545
2	angin pute	284
3	bencana alam	212
-	-	151

no	word	n
-	-	134
-	-	115
7	korban bencana	98

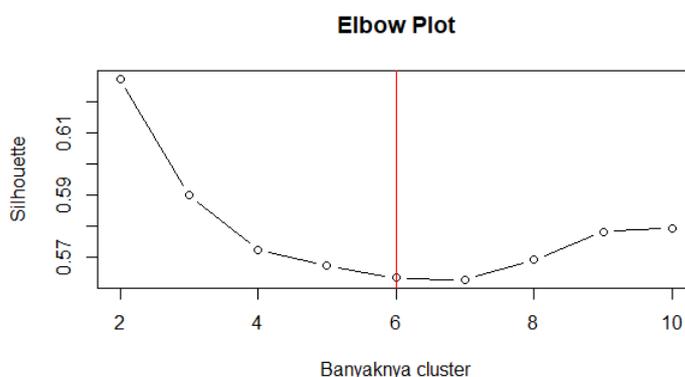
K-Means

Untuk menentukan jumlah *cluster*, digunakan metode *Elbow Plot*. Hasilnya, diperoleh enam klaster pada Gambar 1.



Gambar 2: Elbow Plot

Proses kedua melibatkan perhitungan jarak antar *cluster* menggunakan metode *Silhouette Coefficient* dengan hasil jarak sebesar 0,566. Nilai jarak *cluster* pada Gambar 2.



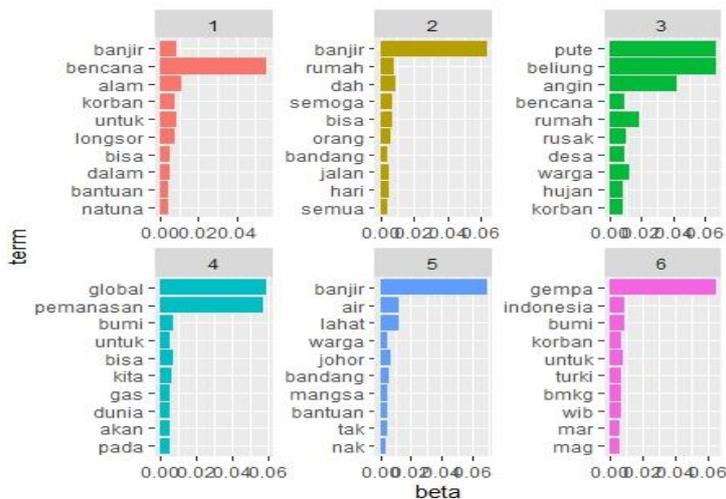
Gambar 3: Silhouette Elbow Plot

Proses ketiga adalah memasukkan data ke masing-masing *cluster* yang ditunjukkan pada Tabel 9.

Tabel 9: Jumlah Anggota Tiap *Cluster*

Cluster	1	2	3	4	5	6
Jumlah Anggota	406	506	506	505	421	575

Tahap keempat melibatkan pengolahan hasil sebelumnya dengan menggunakan metode Linear Discriminant Analysis untuk merepresentasikan k topik. Representasi topik yang dihasilkan dapat dilihat pada Gambar 3.



Gambar 4: Linear Discriminant Analysis

Langkah kelima melibatkan evaluasi *K-Means* dengan menggunakan metode *Sum of Square Error* untuk menghasilkan nilai 97,2%, yang diperoleh dari perhitungan *TF-IDF*. Hasil evaluasi *K-Means* kemudian ditampilkan dalam Gambar 4.

```
within cluster sum of squares by cluster:
[1] 13.727345  8.474946 14.693733 14.776169  8.404964 22.206689
(between_ss / total_ss = 97.2 %)
```

Gambar 5: Hasil Akurasi *K-Means*

Metode *Silhouette Coeffisien* digunakan dengan menggabungkan kluster *Sum of Square Error* dan berhasil mendapatkan akurasi kluster sebesar 96,8%. Hasil evaluasi kluster ditampilkan pada Gambar 5.

```
within cluster sum of squares by cluster:
[1] 8.444941 14.693733 33.442486 4.199988 7.781690 23.555937
(between_ss / total_ss = 96.8 %)
```

Gambar 6: Hasil Akurasi Cluster

Tabel 10 menunjukkan hasil dari penerapan algoritma *K-Means* untuk membuat model pengelompokan bencana alam.

Tabel 10: Hasil Pengelompokan *K-Means*

no	text	Klaster	Kategori_Klaster
1	mum ckp klau batu pahat banjir dia memang lambat turun n lambat surut mak aku ckp pernah jadi bila thn tah n bulan bulan nk tggu air dia surut wey pastu mak aku ckp situ org panggil dia air menung air x gerak tp ahad nie air pasang xtau la tu mcm mana	1	Banjir
2	udh minggu ga workout rumah baru menit udh banjir engapp	1	Banjir
3	urus tuh banjir pati sake	1	Banjir
-	adik aku g kerja yg jalan selalu ambik masa minit jadi jam minit sebab jem	1	Banjir
-	lampau panjang air banjir limpah ke jalan utama kat area bukit pasir tongkang -	1	Banjir
-	-	1	Banjir
-	sementara pps yang masih operasi di enam daerah adalah takat ini	1	Banjir
4169	gempa m di gunungkidul malam ini	6	Gempa

IV. KESIMPULAN

Dalam penelitian ini, Text Mining digunakan untuk membuat model topik tentang bencana alam di Indonesia. Dari hasil penelitian ini, ditemukan bahwa Algoritma K-Means berhasil mengelompokkan topik berdasarkan jarak keterkaitan antar kata-kata. Evaluasi model menggunakan Algoritma K-Means menghasilkan nilai akurasi sebesar 97.2%.

DAFTAR PUSTAKA

- [1] S. M. Hutabarat and A. Sindar, "Data Mining Penjualan Suku Cadang Sepeda Motor Menggunakan Algoritma K-Means," *J. Nas. Komputasi dan Teknol. Inf.*, vol. 2, no. 2, p. 126, 2019, doi: 10.32672/jnkti.v2i2.1555.
- [2] 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, [Online]. Available: <https://t.co/FXtzMcbdHp>
- [3] D. A. Manalu and G. Gunadi, "Implementasi Metode Data Mining K-Means Clustering Terhadap Data Pembayaran Transaksi Menggunakan Bahasa Pemrograman Python Pada Cv Digital Dimensi," *Infotech J. Technol. Inf.*, vol. 8, no. 1, pp. 43–54, 2022, doi: 10.37365/jti.v8i1.131.
- [4] E. Susanto, V. C. Mawardi, and M. D. Lauro, "APLIKASI CLUSTERING BERITA DENGAN METODE K MEANS DAN PERINGKAS BERITA DENGAN METODE MAXIMUM MARGINAL RELEVANCE," *J. Ilmu Komput. dan Sist. Inf.*, vol. 9, no. 1, 2021, doi: 10.24912/jiksi.v9i1.11560.
- [5] M. F. H. Adi and Y. A. Susetyo, "Pemetaan Potensi Bencana di Jawa Tengah Menggunakan Google Maps API dan KML dengan Metode," *Indones. J. Comput. Model.*, 2018.
- [6] Taufiq Al Ashfahani Qodrifuddin *et al.*, "Peningkatan Pemahaman Masyarakat terhadap Bahaya dan Dampak Bencana Alam Serta Penanggulangannya," *J. Pengabd. Magister Pendidik. IPA*, vol. 5, no. 1, 2022, doi: 10.29303/jppmpi.v5i1.1400.
- [7] F. M. Faruk, F. M. Faruk, F. S. Doven, and B. Budyana, "PENERAPAN METODE REGRESI LOGISTIK BINER UNTUK MENGETAHUI DETERMINAN KESIAPSIAGAAN RUMAH TANGGA DALAM MENGHADAPI BENCANA ALAM," *Semin. Nas. Off. Stat.*, vol. 2019, no. 1, 2020, doi: 10.34123/semnasoffstat.v2019i1.146.
- [8] F. A. Alijoyo and Y. Munawar, "Faktor Yang Mempengaruhi Maturitas Manajemen Risiko Organisasi Di Indonesia," *Bina Ekon.*, vol. 23, no. 1, pp. 67–79, 2021, doi: 10.26593/be.v23i1.4366.67-79.
- [9] M. A. Ayu, E. Irawan, and T. Mantoro, "Text mining approaches for analyzing an Indonesian tafseer and translation of the Holy Quran," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 25, no. 3, pp. 1469–1480, 2022, doi: 10.11591/ijeecs.v25.i3.pp1469-1480.
- [10] Y. Sunoto and B. Wasito, "Analisis Testimonial Wisatawan Menggunakan Text Mining Dengan Metode Naive Bayes Dan Decision Tree , Studi Kasus Pada Hotel – Hotel Di Jakarta," *J. Inform. dan Bisnis Anal.*, vol. 3, no. 2, pp. 39–49, 2014.
- [11] I. Olive, D. Putra, K. R. Prilianti, P. Lucky, and T. Irawan, "Implementasi Text Mining untuk Analisis Layanan Transportasi Online dengan Analisis Faktor," *J. SimanteC*, vol. 8, no. 2, pp. 1–9, 2020.
- [12] L. Rahmawati, S. Widya Sihwi, and E. Suryani, "Analisa Clustering Menggunakan Metode K-Means Dan Hierarchical Clustering (Studi Kasus : Dokumen Skripsi Jurusan Kimia, Fmipa, Universitas Sebelas Maret)," *J. Teknol. Inf. ITSmart*, vol. 3, no. 2, p. 66, 2020, doi: 10.20961/its.v3i2.654.
- [13] B. Harahap, "Penerapan Algoritma K-Means Untuk Menentukan Bahan Bangunan Laris (Studi Kasus Pada UD. Toko Bangunan YD Indarung)," pp. 394–403, 2019.
- [14] S. J. Patandianan, G. M. Putra, and H. S. Pakpahan, "Implementasi Metode K-Means untuk Pengelompokan Rekomendasi Tugas Akhir," vol. 16, no. 1, 2021.
- [15] S. M. Fani, R. Santoso, and S. Suparti, "Penerapan Text Mining Untuk Melakukan Clustering Data Tweet Akun Blibli Pada Media Sosial Twitter Menggunakan K-Means Clustering," *J. Gaussian*, vol. 10, no. 4, pp. 583–593, 2021, doi: 10.14710/j.gauss.v10i4.30409.