

CLUSTERING K-MEANS UNTUK ANALISIS POLA PERSEBARAN BENCANA ALAM DI INDONESIA

by Agus Rahmad Hidayah

Submission date: 14-Jul-2023 03:04PM (UTC+0700)

Submission ID: 2120139944

File name: K_ANALISIS_POLA_PERSEBARAN_BENCANA_ALAM_DI_INDONESIA_108-114.pdf (800.92K)

Word count: 1924

Character count: 11731



JURNAL INFORMATIKA DAN TEKNOLOGI KOMPUTER

Halaman Jurnal: <https://journal.amikveteran.ac.id/index.php/jitek>

Halaman UTAMA Jurnal: <https://journal.amikveteran.ac.id/index.php>



DOI : <https://doi.org/10.55606/jitek.v3i2.1506>

CLUSTERING K-MEANS UNTUK ANALISIS POLA PERSEBARAN BENCANA ALAM DI INDONESIA

M Aditya Yoga Pratama^a, Agus Rahmad Hidayah^{b*}, Tertia Avini^c

^a Ilmu Komputer / Teknik Informatika, 2021110023@students.uigm.ac.id, Universitas Indo Global Mandiri

^b Ilmu Komputer / Teknik Informatika, 2021110021@students.uigm.ac.id, Universitas Indo Global Mandiri

^c Ilmu Komputer / Sistem Operasi, avini.saputra@uigm.ac.id, Universitas Indo Global Mandiri

* Correspondence

ABSTRACT

Data clustering plays a crucial role in data analysis for identifying hidden patterns, trends, and structures within the data. The K-Means algorithm has gained popularity as a widely used method for data clustering due to its efficiency and ease of implementation. Clustering is a data analysis technique utilized to group similar objects together. The K-Means algorithm stands out as one of the most renowned and frequently employed clustering methods across various fields, including data science, pattern recognition, and artificial intelligence. In this research, we collected data on natural disasters from different regions in Indonesia and employed it as input for the K-Means clustering algorithm. K-Means was utilized to cluster the similarity patterns within the occurring natural disasters. The clustering results provide information about groups that may exhibit similar characteristics and disaster risks.

Keywords: Clustering, K-means, Natural Disasters, Algorithms, Data Mining

Abstrak

Pengelompokan data memiliki peranan penting dalam analisis data untuk mengidentifikasi pola tersembunyi, tren, dan struktur di dalamnya. Algoritma K-Means telah menjadi metode yang populer dan banyak digunakan dalam pengelompokan data karena efisiensinya dan kemudahan implementasinya. Pengelompokan merupakan teknik analisis data yang digunakan untuk mengelompokkan objek yang memiliki kesamaan. Algoritma K-Means menjadi salah satu metode pengelompokan yang paling terkenal dan sering digunakan di berbagai bidang, termasuk ilmu data, pengenalan pola, dan kecerdasan buatan. Dalam penelitian ini, kami mengumpulkan data tentang bencana alam dari berbagai wilayah di Indonesia dan menggunakan data tersebut sebagai input untuk algoritma pengelompokan K-Means. K-Means digunakan untuk mengelompokkan pola-pola kesamaan dalam bencana alam yang terjadi. Hasil pengelompokan memberikan informasi tentang kelompok-kelompok yang mungkin memiliki karakteristik dan risiko bencana yang serupa..

Kata Kunci: Clustering, K-means, Bencana Alam, Algoritma, Data Mining

1. PENDAHULUAN

Seiring berkembangnya Ilmu teknologi memudahkan kita dalam mendapatkan data dan informasi saat mengakses internet. data dari internet pastilah banyak dan berlimpah saat kita memberikan pertanyaan pada mesin pencarian. Tetapi apakah data yang kita akses akan selalu benar dan tepat[1]. Oleh karena itu, Metode Clustering merupakan suatu pengelompokan data yang dapat menjadi beberapa kluster sehingga dapat memberikan data yang relevan dan terstruktur[2]. Ada dua tipe data clustering yaitu diantaranya ada Hierarchical (hirarki) dan non Hierarchical (non- hirarki) yang dimana dua tipe data tersebut tergolong dalam Data Mining yang bersifat Unsupervised (tanpa arahan). K-Means adalah metode clustering non

Received Mei 23, 2023; Revised Juni 21, 2023; Accepted Juli 15, 2023

hirarki yang mempartisi data mengolah dan membent⁶ satu atau lebih klaster [3]. Indonesia merupakan wilayah kawasan Asia Tenggara yang memiliki letak geologis, de⁸ grafis, dan geografis yang memicu terjadinya bencana alam yang seringkali muncul seperti contohnya gempa bumi, banjir, gunung Meletus, angin putting beliung dan sebagainya [4]. Tujuan penulisan jurnal ini guna mengetahui pengelompokan suatu data pada bencana alam yang terjadi di wilayah Indonesia, menggunakan algoritma K-Means pada aplikasi software Rapidminer guna memudahkan suatu data dikelompokkan sehingga lebih efisien dan terstruktur.

2. METODOLOGI PENELITIAN

Metode Penelitian Kuantitatif

Metode ini menggunakan pendekatan kuantitatif dalam pengumpulan dan analisis data. Data yang dikumpulkan dalam bentuk angka dan diolah menggunakan analisis statistik untuk menghasilkan temuan.

Metode clustering

Metode clustering pada algoritma K-means adalah teknik pengelompokan data yang populer dan efisien, langkah-langkahnya

- Inisialisasi: Langkah pertama adalah menentukan jumlah kelompok yang diinginkan (K) dan menginisialisasi posisi awal dari pusat kelompok secara acak. Pusat kelompok awal ini disebut centroid.
- Pengelompokan: Setiap titik data dalam dataset akan dikelompokkan ke dalam kelompok yang memiliki centroid terdekat. Jarak antara titik data dengan centroid diukur menggunakan metrik jarak Euclidean atau metrik jarak lainnya.
- Perbarui centroid: Setelah pengelompokan awal, centroid baru dihitung dengan mengambil rata-rata dari semua titik data dalam kelompok. Posisi centroid diperbarui untuk setiap kelompok.
- Iterasi: Langkah pengelompokan dan pembaruan centroid diulang secara iteratif. Pada setiap iterasi, titik data diklasifikasikan ke kelompok yang memiliki centroid terdekat, dan centroid diperbarui berdasarkan rata-rata kelompok.
- Konvergensi: Iterasi dilanjutkan hingga tidak ada perubahan dalam pengelompokan atau pusat kelompok lagi. Pada titik ini, algoritma dianggap telah mencapai konvergensi, dan hasil akhir adalah kelompok-kelompok yang stabil.
- Evaluasi: Setelah konvergensi, hasil clustering dapat dievaluasi menggunakan metrik evaluasi seperti SSE (Sum of Squared Errors) atau indeks kecocokan yang sesuai dengan tujuan penelitian.

Berikut kami berikan Langkah-langkahnya dalam mengelompokkan data nya pada aplikasi software RapidMining :

- Pertama kita siapkan dulu data mentahnya pada software Ms Excel

Provinsi	Tanah Longsor	Banjir	Banjir Bandang	Gempa Bumi	Tsunami	Gelombang Pasang Laut
ACEH	198	1435	81	493	2	106
SUMATERA UTARA	483	732	52	364	4	76
SUMATERA BARAT	222	342	65	364	0	57
RIAU	21	495	1	0	0	15
JAMBI	57	476	17	36	0	2
SUMATERA SELATAN	103	380	36	49	0	3
BENGKULU	81	171	15	56	0	14
LAMPUNG	70	328	23	47	0	35
KEP. BANGKA BELITUNG	1	59	0	0	0	17
KEP. RIAU	25	61	1	0	0	69
DKI JAKARTA	7	109	0	1	0	5
JAVA BARAT	1288	1933	300	601	0	91
JAVA TENGAH	1190	1249	69	114	0	128
DIYOGYAKARTA	59	50	1	52	0	12
JAVA TIMUR	511	1176	36	2449	4	67
BANTEN	36	419	41	158	1	39
BALI	105	39	5	79	0	34
NUSA TENGGARA BARAT	44	187	54	100	0	44
NUSA TENGGARA TIMUR	610	570	151	246	0	218
KALIMANTAN BARAT	66	884	22	0	0	48
KALIMANTAN TENGAH	24	735	9	0	0	12
KALIMANTAN SELATAN	74	329	31	0	0	30
KALIMANTAN TIMUR	74	343	5	6	0	9
KALIMANTAN UTARA	42	200	0	18	0	3
SULAWESI UTARA	218	288	37	349	0	95
SULAWESI TENGAH	133	587	57	558	0	117
SULAWESI SELATAN	251	620	52	272	0	39
SULAWESI TENGGARA	46	240	10	43	0	58
GORONTALO	65	259	24	95	0	37
SULAWESI BARAT	144	125	7	498	0	35
MALUKU	58	153	2	106	0	105
MALUKU UTARA	80	270	20	569	0	124
PAPUA BARAT	72	138	0	318	0	30
PAPUA	146	204	9	57	0	57
INDONESIA	6664	15366	1093	6726	11	1913

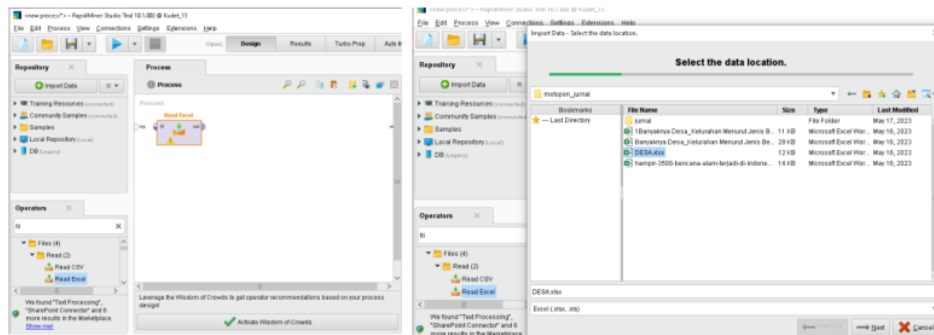
Angin Puting Belulang	Gunung Meletus	Kebakaran Hutan	Kekeringan	Tidak Ada Bencana Alam
108	1	43	173	4406
483	62	59	127	3827
248	0	18	43	513
53	0	154	51	1224
44	0	16	16	1001
30	0	64	98	2644
17	0	4	19	1195
158	0	11	30	2093
77	0	14	0	265
40	0	57	27	241
0	0	1	0	152
412	0	76	442	3081
619	26	49	278	5637
84	3	4	31	233
486	74	50	156	4655
107	0	8	77	314
46	0	0	2	482
65	0	19	88	748
1219	54	144	454	1487
53	0	135	75	1158
16	0	66	37	789
70	0	60	35	373
26	0	52	32	628
3	0	13	6	248
107	0	25	24	1075
85	0	29	42	984
233	0	22	60	1863
104	0	12	16	1852
34	0	9	57	352
26	0	8	22	105
54	0	26	26	798
42	1	16	12	423
5	0	12	5	1482
11	0	14	9	5142
5206	241	1338	2570	52670

3

Gambar 2.1. Data Mentah Bencana Alam pada tahun 2021

Sumber : <https://www.bps.go.id/indicator/168/954/1/banyaknya-desa-kelurahan-menurut-jenis-bencana-alam-dalam-tiga-tahun-terakhir.html>

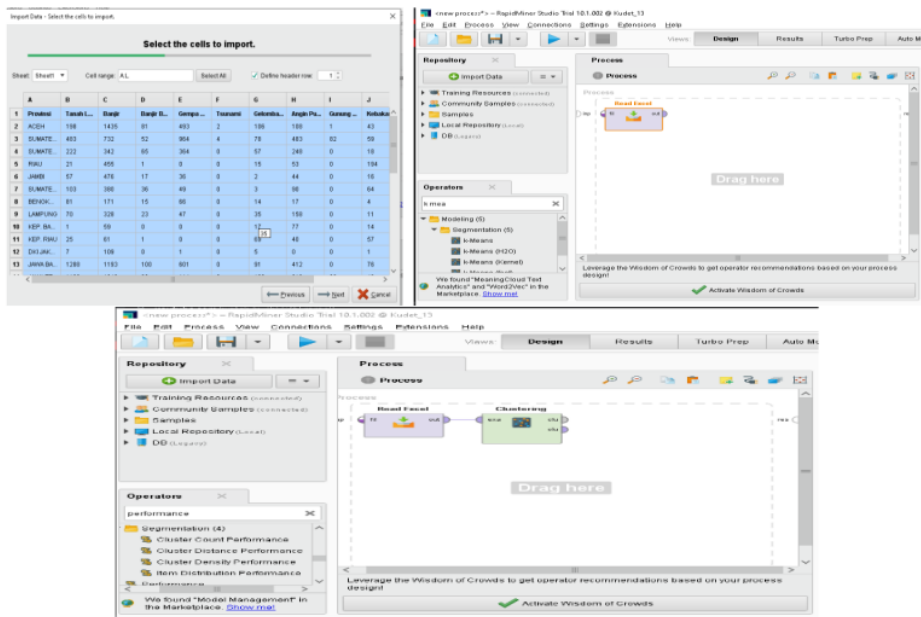
- b) Kemudian rapikan datanya hilangkan caption bencananya dan provinsi
- c) Lalu, masuk ke software RapidMining klik file dan new file, kemudian ketikan di operator “excel” sesuai format file kalian lalu drag file ke layer



Gambar 2.2. Tampilan add new document, dan memasukkan file data mentah pada RapidMining

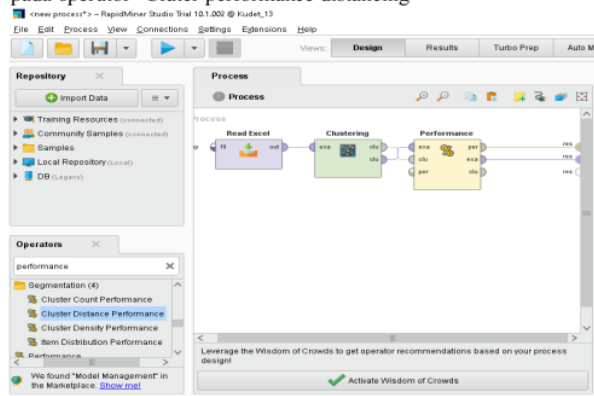
- d) Kemudian import data cell, ubah semua data menjadi integer, kemudian pilih K-Means pada operator dibawah

1



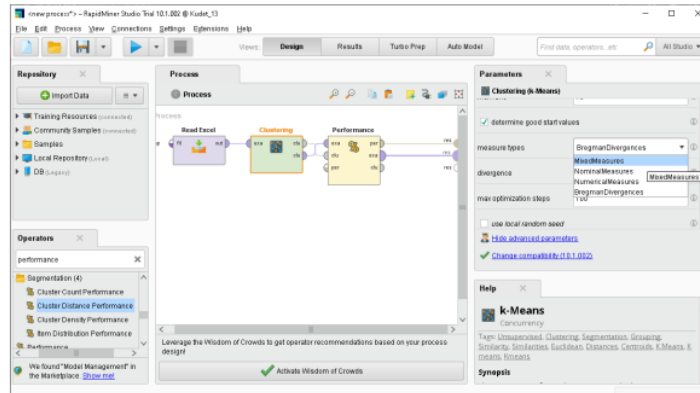
Gambar 2.3. Import cell dan operator K-means

e) Kemudian add pada operator “Cluter performance distancing”



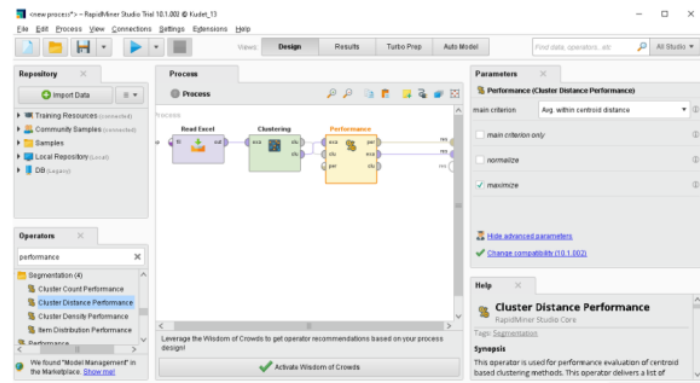
Gambar 2.4. add cluter performance distancing

f) Setelah itu klik pada bagian cluterling K-Means dan ubah parameternya di bagian measure types nya menjadi “mixedmeasures” agar nilai selain Integer dapat direading



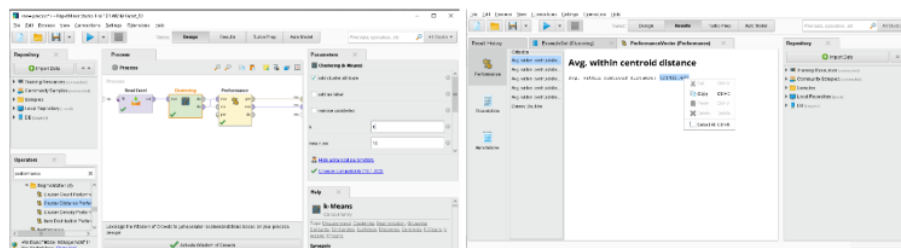
Gambar 2.5. mengubah parameter

- g) Pada bagian Cluster Distance Performes ceklist menjadi Maximize agar tidak ada nilai selain nilai positif



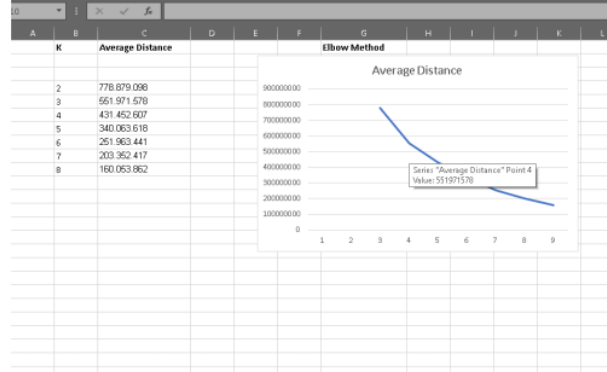
Gambar 2.6. mengubah nilai Maximize

- h) Pada Clustering K-Means dipojok bagian sebelah kanan pada "K" diubah satu persatu dari "K" 2 – 8. Dan lihat Average centroid distance nya pada setiap K.



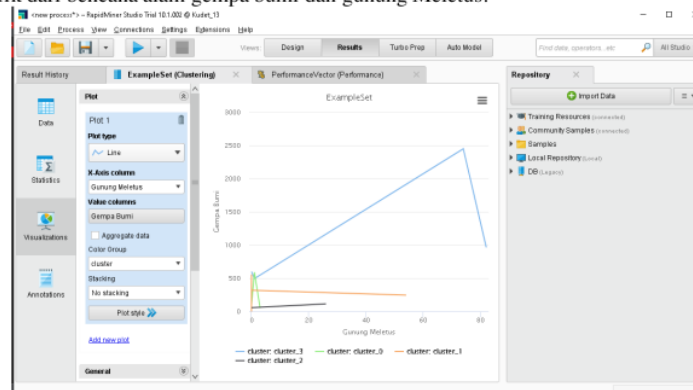
Gambar 2.7. mengubah setiap nilai Average dari K

- i) Jika semua nilai “K” sudah di “Average” lalu di blok dan klik kanan pilih copy setiap nilai Average centroid distance dan masuk Kembali ke Excel.
- j) Jika sudah masuk ke Excel lalu kita masukkan data dari “K” 2 – 8 beserta Averagennya



Gambar 2.8. Menginputkan semua nilai

- k) Seperti pada gambar 2.8 diatas bahwa untuk mencari “K” untuk Cluter yang baik maka kita disini menggunakan Teknik Elbow Method dimana, Elbow Method ini mencari suatu nilai K dari sebuah sudut yang terbilang seperti siku. Disini dapat dilihat bahwa sudut siku tersebut berada di angka 4 pada grafik kurva, jadi nilai yang baik untuk Cluter nya dapat kita isi dengan nilai yang sudah didapat “4”.
- l) Dan inilah hasil nya dapat kita lihat dari cluternya yang telah memiliki “4”, kemudian disitu ada Data dan Grafik dari bencana alam gempa bumi dan gunung Meletus.



Gambar 2.9. Contoh Hasil pada Bencana alam Gempa dan Gunung Meletus

3. HASIL DAN PEMBAHASAN

Berikut adalah simpulan dari penelitian yang telah kami implementasikan menggunakan aplikasi software RapidMining. Dalam era internet, akses mudah terhadap data dan informasi menjadi lebih memungkinkan, namun keakuratan data tetap menjadi tantangan yang perlu diatasi. Oleh karena itu, metode clustering dapat digunakan untuk mengelompokkan data secara efisien sehingga memfasilitasi akses terhadap data yang relevan dan terstruktur.

Metode clustering terdiri dari beberapa tipe, termasuk hierarchical dan non-hierarchical, dan salah satu metode non-hierarchical yang populer adalah K-Means². Di Indonesia, bencana alam sering terjadi, sehingga pengelompokan data bencana alam menggunakan metode K-Means dapat membantu mengorganisir data secara efisien dan terstruktur. Penelitian ini menggunakan aplikasi software RapidMining untuk menerapkan algoritma K-Means pada data bencana alam di Indonesia.

Diharapkan penggunaan metode clustering ini dapat memberikan manfaat dalam mengelompokkan data yang relevan dan terstruktur, serta membantu dalam mitigasi dan penanganan bencana alam di Indonesia.

4. KESIMPULAN DAN SARAN

Berikut yang dapat kami simpulkan mengenai penelitian yang sudah kami Implementasikan dalam sebuah bentuk Aplikasi Software RapidMining Penggunaan internet telah memudahkan akses terhadap data dan informasi, namun keakuratan data yang didapatkan masih menjadi masalah yang perlu diatasi. Oleh karena itu, metode clustering dapat digunakan untuk mengelompokkan data sehingga memudahkan dalam mengakses data yang relevan dan terstruktur.

Metode clustering terdiri dari dua tipe, yaitu hierarchical dan non-hierarchical, dan K-Means merupakan salah satu metode clustering non-hierarchical yang populer. Di Indonesia⁴, bencana alam sering terjadi, sehingga pengelompokan data bencana alam dengan menggunakan metode K-Means dapat membantu dalam mengelompokkan data dengan efisien dan terstruktur. Penelitian ini menggunakan aplikasi software Rapidminer untuk menerapkan algoritma K-Means pada data bencana alam di Indonesia. Diharapkan, penggunaan metode clustering dapat memberikan manfaat dalam mengelompokkan data yang relevan dan terstruktur, serta dapat membantu dalam mitigasi dan penanganan bencana alam di Indonesia.

5. DAFTAR PUSTAKA

- [1] K. Jacksi and N. Salih, "State of the art document clustering algorithms based on semantic similarity," *J. Inform.*, vol. 14, no. 2, p. 58, 2020, doi: 10.26555/jifo.v14i2.a17513.
- [2] A. F. Khairati, A. . Adlina, G. . Hertono, and B. . Handari, "Kajian Indeks Validitas pada Algoritma K-Means Enhanced dan K-Means MMCA," *Prism. Pros. Semin. Nas. Mat.*, vol. 2, pp. 161–170, 2019, [Online]. Available: <https://journal.unnes.ac.id/sju/index.php/prisma/article/view/28906>
- [3] Y. Agusta, "K-Means-Penerapan, Permasalahan dan Metode Terkait," *J. Sist. dan Inform.*, vol. 3, no. Pebruari, pp. 47–60, 2007.
- [4] D. I. Ramadhani, O. Damayanti, O. Thaushiyah, and A. R. Kadafi, "Penerapan Metode K-Means Untuk Clustering Desa Rawan Bencana Berdasarkan Data Kejadian Terjadinya Bencana Alam," *JURIKOM (Jurnal Ris. Komputer)*, vol. 9, no. 3, p. 749, 2022, doi: 10.30865/jurikom.v9i3.4326.
- [5] T. Terttiaavini and T. S. Saputra, "Analisa Akurasi Penggunaan Metode Single Eksponential Smoothing untuk Perkiraan Penerimaan Mahasiswa Baru Pada Perguruan Tinggi XYZ," *J. Inform. Glob.*, vol. 11, no. 1, 2020.
- [6] F. Z. Terttiaavini, M. Ramadhan, K. Azrai'ie, and T. S. Saputra, "Clustering Analysis of Premier Research Fields," 2016.
- [7] E. Berliana, N. Nurhalizah, N. Wahyuni, L. Hertati, T. Terttiaavini, and A. Asmawati, "Peran Dunia Digital Sistem Informasi Manajemen Pembelian Online Dan Offline Yang Bekerja Melayani Konsumen," *Glow J. Pengabd. Kpd. Masy.*, vol. 2, no. 2, pp. 44–55, 2022.
- [8] N. Suhandi and R. G. Terttiaavini, "Decision Support System for Subsidized Housing Selection Based on Best-Worst Method and Simple Additive Weighting," *Int. J.*, vol. 9, no. 3, 2020.
- [9] A. Sanmorino, R. Gustriansyah, T. Terttiaavini, and I. Isabella, "The Toolkit of Success Rate Calculation of Broiler Harvest," *TELKOMNIKA (Telecommunication Comput. Electron. Control.*, vol. 15, no. 4, pp. 1947–1954, 2017.

CLUSTERING K-MEANS UNTUK ANALISIS POLA PERSEBARAN BENCANA ALAM DI INDONESIA

ORIGINALITY REPORT

14%

SIMILARITY INDEX

14%

INTERNET SOURCES

7%

PUBLICATIONS

4%

STUDENT PAPERS

PRIMARY SOURCES

1

journal.amikveteran.ac.id

Internet Source

7%

2

eprints.umpo.ac.id

Internet Source

2%

3

lintangsegoro.com

Internet Source

1%

4

docobook.com

Internet Source

1%

5

jurnal.stmikroyal.ac.id

Internet Source

1%

6

ejurnal.stmik-budidarma.ac.id

Internet Source

<1%

7

id.scribd.com

Internet Source

<1%

8

repository.unj.ac.id

Internet Source

<1%

9

www.in2connect.uk.com

Internet Source

<1%

Exclude quotes Off

Exclude matches Off

Exclude bibliography Off