

TINJAUAN PUSTAKA SISTEMATIS PADA DATA MINING: STUDI KASUS ALGORITMA K-MEANS CLUSTERING

Sekar Setyaningtyas¹⁾, Bangkit Indarmawan Nugroho²⁾, Zaenul Arif³⁾

^{1,3}Program Studi Teknik Informatika

²Program Studi Sistem Informasi

^{1,2,3}Sekolah Tinggi Manajemen Informatika dan Komputer Tegal

E-mail : sekarsetya240@gmail.com ¹⁾, efbeterang@gmail.com ²⁾, zendhunter@gmail.com ³⁾

Abstract

Data Mining is a method for analyzing future patterns and characteristics as well as gathering unexpected, never-before-seen information from large databases. In data mining, clustering is one of the useful techniques for analyzing data. One of the data mining algorithms is the K-Means algorithm, which is a clustering technique based on distance division. The goal to be achieved in this paper is to analyze the clustering technique using the K-Means algorithm in data mining by conducting an in-depth review and searching through the literature selected based on the criteria and studies that will be selected to answer research questions. Systematic Literature Review (SLR) is a method that aims to identify and find research results with techniques based on specific procedures from comparison results. Based on the literature on the selection of journal publications, Pattern Recognition, Knowledge-Based Systems, Applied Soft Computing and IEEE Access can be the main references related to the K-Means algorithm. The results of the comparison show that Euclidean Distance has the advantage of better distance calculation, so that this method can be used as the main choice related to the calculation theory of the K-Means algorithm.

Keywords- data mining, clustering, K-Means, systematic literature review.

Intisari

Data Mining adalah metode untuk menganalisis pola dan karakteristik di masa depan serta untuk mengumpulkan informasi tak terduga yang belum pernah terlihat sebelumnya dari database yang besar. Dalam data mining, clustering adalah salah satu teknik yang berguna untuk analisis data. Salah satu algoritma data mining adalah algoritma K-Means yang merupakan teknik clustering berdasarkan pembagian jarak. Tujuan yang ingin dicapai dalam paper ini yakni menganalisis teknik clustering menggunakan algoritma K-Means dalam data mining dengan melakukan review secara mendalam dan mengevaluasi penelusuran melalui literatur terpilih berdasarkan kriteria tertentu dan studi yang dipilih akan diproses untuk menjawab pertanyaan penelitian. Tinjauan Pustaka Sistematis (Systematic Literature Review/SLR) merupakan sebuah metode penelitian yang bertujuan untuk mengidentifikasi dan mengevaluasi hasil penelitian dengan teknik terbaik berdasarkan prosedur yang spesifik dari hasil perbandingan. Berdasarkan pemilihan literatur publikasi jurnal, Pattern Recognition, Knowledge-Based System, Applied Soft Computing dan IEEE Access dapat menjadi rujukan utama terkait algoritma K-Means. Hasil perbandingan metode menunjukkan bahwa Euclidean Distance memiliki keunggulan perhitungan jarak yang lebih baik, sehingga metode tersebut dapat dijadikan sebagai pilihan utama terkait teori perhitungan jarak pada algoritma K-Means.

Kata Kunci—data mining, clustering, K-Means, tinjauan pustaka sistematis.

1. PENDAHULUAN

Data Mining adalah metode untuk menganalisis pola dan karakteristik di masa depan serta untuk mengumpulkan informasi tak terduga yang belum pernah terlihat sebelumnya dari database yang besar [1]. Data Mining mengeksplorasi pengetahuan dan pola dalam data melalui statistik matematika dan *machine learning* [2].

Penggunaan sistem informasi terdistribusi sangat berpengaruh dalam menyebabkan berkembangnya sebuah kumpulan data menjadi data yang sangat besar. Seiring berjalannya waktu, suatu data akan terus bertambah dari data yang sekarang digabungkan dengan data di masa depan sehingga akan ada aliran data yang besar, algoritma data mining berfungsi secara efektif dan efisien untuk menganalisis data yang besar [3].

Pada data mining, *clustering* merupakan suatu teknik untuk menganalisis data [4]. Tujuan dari *clustering* adalah untuk mempartisi data menjadi suatu kelompok [5]. *Clustering* mengelompokkan kumpulan data ke dalam kelompok-kelompok yang memiliki kesamaan ke dalam *cluster* yang sama dan yang tidak sama ke *cluster* yang sama [4]. Metode *clustering* telah banyak digunakan di berbagai bidang seperti bidang penelitian dan pemodelan identifikasi objek, pengolahan citra dan pemetaan zona wilayah [6].

Teknik pengelompokan paling populer dalam bidang ilmiah salah satunya adalah algoritma K-Means [7]. Algoritma K-Means adalah salah satu algoritma dengan teknik *clustering* berdasarkan pembagian jarak dalam data mining. Keuntungan menggunakan algoritma ini mudah dipahami, diterapkan, serta memiliki efek pengelompokan yang baik sehingga K-Means banyak digunakan dalam bidang penelitian [8]. Namun, algoritma ini juga memiliki kekurangan, yakni memiliki ketergantungan yang kuat pada pemilihan pusat *cluster* awal [9].

Dengan menerapkan *Systematic Literature Review*, tujuan yang ingin dicapai dalam paper ini untuk menganalisis data mining dengan menggunakan teknik *clustering* dalam algoritma K-Means dengan melakukan *review* secara mendalam dan mengevaluasi penelusuran melalui literatur terpilih menggunakan parameter tertentu dan subjek yang terpilih akan diproses untuk menjawab pertanyaan penelitian.

Penelitian dengan menggunakan *Systematic Literature Review* dalam studi data mining teknik *clustering* algoritma K-Means telah dilakukan oleh peneliti sebelumnya, [10] dan [11]. Dalam penelitian tersebut menghasilkan beberapa paper yang kemudian di *review* secara rinci sehingga menghasilkan evaluasi literatur yang dapat menjawab pertanyaan penelitian.

Pada penelitian terkait *Systematic Literature Review* yang dipublikasikan oleh [10], menganalisis beberapa paper terkait Penggunaan algoritma K-Means Untuk *Clustering* Bidang Pendidikan dengan jangka waktu 2017-2021 yang didapatkan melalui pencarian dari beberapa perpustakaan digital. Hasil yang diperoleh adalah 12 paper final

yang dilakukan setelah evaluasi penelusuran literatur secara mendalam.

Menurut penelitian [11] yang dilakukan pada jangka waktu 2015-2020, *Clustering* K-Means masih dianggap sebagai salah satu teknik data mining yang banyak digunakan. Dari penelitian tersebut, dengan menggunakan *Systematic Literature Review* telah ditinjau hasil studi tentang *Clustering* K-Means berdasarkan kriteria pencarian dan ditemukan 835 paper dari 5 perpustakaan digital. Jumlah penelitian dengan menggunakan teknik ini sangat signifikan menunjukkan bahwa meskipun K-Means telah digunakan secara luas di berbagai bidang, tetapi K-Means masih menjadi salah satu teknik paling populer diantara penelitian data mining. Sehingga 114 paper berdasarkan judul paper dipilih untuk dianalisis lebih lanjut dan dilakukan *review* secara mendalam.

Dalam penelitian ini, objek utama yang digunakan adalah algoritma K-Means *clustering*, sehingga metode pencarian paper dilakukan berdasarkan dari suatu riset penelitian dengan menggunakan algoritma K-Means *clustering* yang mencakup ke dalam suatu bidang/topik apapun. Metode pencarian paper dilakukan menggunakan *keyword* tertentu dari 3 perpustakaan digital yang masih populer selama 5 tahun terakhir, yakni pada tahun 2017-2021. Penelitian ini hanya membatasi untuk paper dari jurnal yang selanjutnya akan dikaji sehingga menghasilkan penelitian terbaru terkait data mining teknik *clustering* algoritma K-Means.

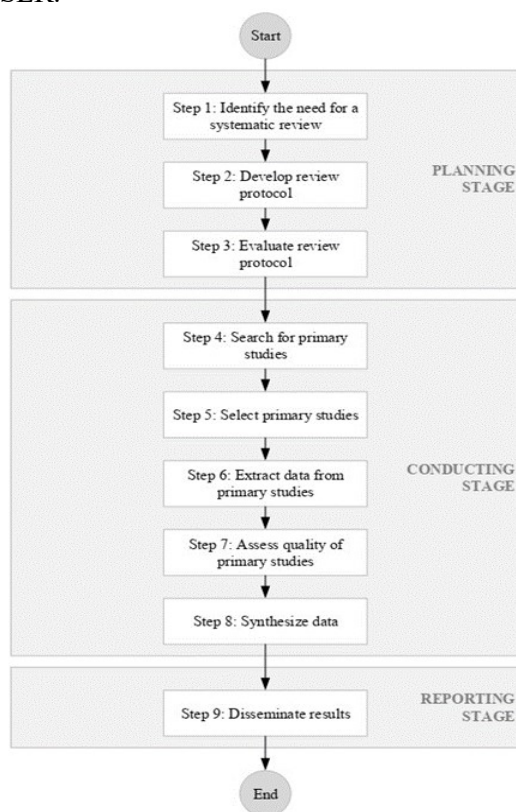
2. METODOLOGI

2.1 Systematic Literature Review

Tinjauan Pustaka Sistematis (*Systematic Literature Review/SLR*) adalah sebuah metode penelitian yang bertujuan untuk mengidentifikasi dan mengevaluasi hasil penelitian dengan teknik terbaik berdasarkan prosedur yang spesifik dari hasil perbandingan [12]. Pada penelitian ini, langkah-langkah yang digunakan mengikuti kaidah penelitian serta mengadopsi prosedur dari penelitian [13].

Systematic Literature Review diproses dalam tiga langkah, yakni: perencanaan, pelaksanaan, dan pelaporan tinjauan pustaka.

Gambar 1. menunjukkan tahapan pada metode SLR.



Gambar 1. Langkah SLR [14]

2.2 Pertanyaan Penelitian

Pertanyaan Penelitian (*Research Question*\RQ) merupakan identifikasi pertama dalam metode SLR. RQ memiliki tujuan yakni melakukan teknik pencarian serta meneliti literatur yang mendalam secara otomatis dengan berorientasi pada objek tujuan. Untuk merancang RQ yang baik maka digunakan pendekatan PICOC [14].

Tabel 1. menyajikan ringkasan untuk RQ dari PICOC. Tabel 2. menyajikan pertanyaan penelitian yang bertujuan untuk membahas lebih lanjut terkait penelitian ini.

Tabel 1. Ringkasan PICOC

Population	Data Mining, <i>Clustering</i> , K-Means
Intervention	Karakteristik K-Means
Comparison	-
Outcomes	Bidang yang meliputi algoritma K-Means secara khusus
Context	Studi khusus terkait algoritma K-Means

Tabel 2. Pertanyaan Penelitian dan Tujuan

ID	Pertanyaan Penelitian	Tujuan
RQ1	Jurnal manakah yang paling aktif mempublikasikan algoritma K-Means?	Mengidentifikasi jurnal yang sering mempublikasikan algoritma K-Means
RQ2	Metode apa yang paling sering digunakan pada algoritma K-Means?	Mengidentifikasi teori perhitungan jarak dalam algoritma K-Means
RQ3	Bagaimana hasil perbandingan metode dalam algoritma K-Means?	Mengidentifikasi perbandingan metode dalam algoritma K-Means

2.3 Strategi Pencarian

Tinjauan pustaka memiliki beberapa bagian tertentu pada proses pencarian, yakni harus menentukan *digital library*, memilih *keyword* tertentu, menerapkan *keyword* yang sudah ditentukan, mengoreksi *keyword* dan mengumpulkan beberapa literatur dari *digital library*. Sebelum memulai pencarian, tahap awal yang harus dilakukan adalah melakukan penentuan atau pemilihan *database* yang sesuai untuk menemukan paper yang relevan. *Database* dari *digital library* yang digunakan adalah Science Direct, Google Scholar, dan IEEE.

Menurut [13], pemilihan pencarian menggunakan *keyword* tertentu dilakukan beberapa langkah sebagai berikut:

1. Identifikasi pencarian berdasarkan kata kunci dengan menggunakan ringkasan PICOC terutama pada *Population* dan *Intervention*.
2. Identifikasi pencarian menggunakan pertanyaan penelitian.
3. Identifikasi pencarian kata kunci menggunakan abstrak dan *keyword* serta judul yang selaras.
4. Identifikasi persamaan kata, lawan kata, dan pergantian kata berdasarkan kata kunci pencarian.
5. Menggunakan kata kunci sebagai pencarian lanjutan dengan menerapkan Boolean AND dan OR.

Kata kunci yang digunakan untuk pencarian:

(Data Mining OR DM*) AND (Clustering OR Cluster*) AND (K-Means*)

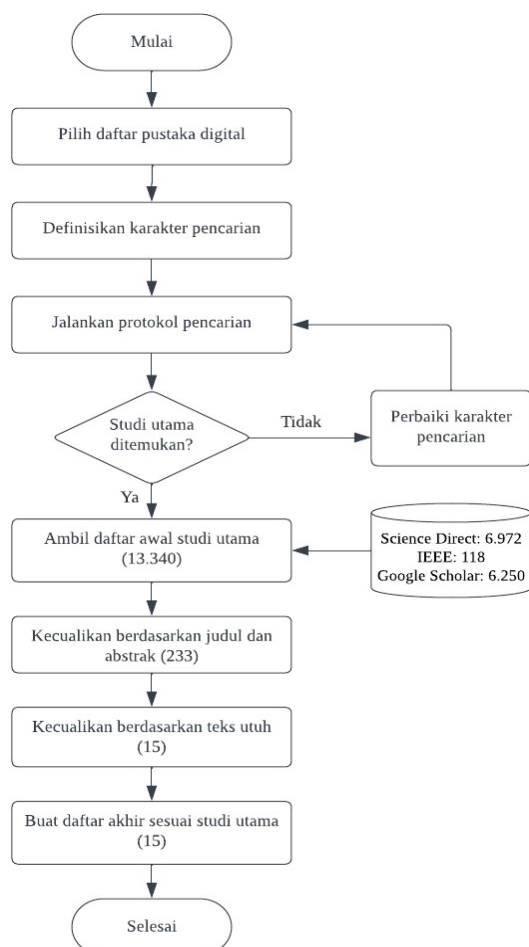
2.4 Penyeleksian Kajian

Rangkuman kriteria inklusi dan kriteria eksklusi digunakan dengan tujuan untuk mendapatkan literatur *primer* yang kemudian akan ditinjau. Tabel 3. menyajikan rangkuman kriteria yang akan dikaji.

Tabel 3. Rangkuman Kriteria

Kriteria Inklusi	Kajian termasuk ke dalam Data Science atau Computer Science
	Diutamakan berbahasa Inggris
Kriteria Eksklusi	Kajian tidak menggunakan validasi
	Kajian yang dilakukan dibawah tahun 2015

Gambar 2. menunjukkan *flowchart* hasil pencarian dan seleksi kajian setelah dikaji.



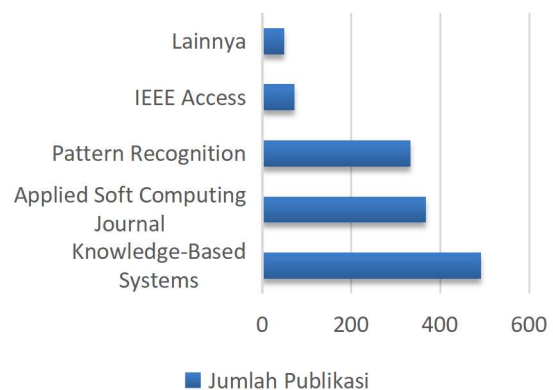
Gambar 2. Hasil Pencarian dan Seleksi Kajian

3. HASIL DAN PEMBAHASAN

3.1 Publikasi Jurnal Berpengaruh

Berdasarkan dari literatur *primer* yang terpilih, publikasi jurnal paling aktif yang mempublikasikan paper terkait algoritma K-Means adalah *Knowledge-Based Systems* sebanyak 491 paper, *Applied Soft Computing Journal* sebanyak 367 paper, *Pattern Recognition* sebanyak 333 paper, *IEEE Access* sebanyak 72 paper. Gambar 3 menyajikan statistik distribusi publikasi artikel dari literatur terpilih.

Jumlah Publikasi



Gambar 3. Distribusi Publikasi Paper dari Literatur Terpilih

Pada tabel 4. menunjukkan kualitas serta ranking publikasi jurnal yang paling aktif. *Pattern Recognition* memiliki ranking tertinggi di *Scimago Journal Rank* (SJR) dan memiliki kualitas yang terbaik serta memiliki jumlah artikel yang cukup banyak dalam mempublikasikan artikel terkait algoritma K-Means.

Tabel 4. Literatur Terpilih serta SJR

Nama Jurnal	SJR	Kategori Q	Database
Pattern Recognition	3.11	Q1	Science Direct
Knowledge-Based Systems	2.19	Q1	Science Direct
Applied Soft Computing Journal	1.96	Q1	Science Direct
IEEE Access	0.92	Q1	IEEE

3.2 Evaluasi Literatur dan Perbandingan Metode

Setelah dilakukan tahap pemilihan literatur dengan berdasarkan jurnal yang paling aktif dalam mempublikasikan paper terkait algoritma K-Means dengan kualitas terbaik, maka tahap selanjutnya adalah mengevaluasi literatur berdasarkan karakteristik melalui

metode serta melakukan perbandingan *database* yang paling sering mempublikasikan terkait algoritma K-Means. Pada tahap ini diperoleh 15 paper final yang memenuhi syarat dari kriteria inklusi dan kriteria eksklusi.

Tabel 5. menyajikan rangkuman evaluasi literatur disertai dengan database yang digunakan.

Tabel 5. Daftar Paper Literatur Terpilih dan Perbandingan Metode serta Database

No	Referensi	Judul	Topik	Metode	Hasil	Jurnal	Database
1	Marco Capo dkk. (2017) [15].	An efficient approximation to the K-Means clustering for massive data	General	Lloyd Algorithm	Terdapat strategi pengurangan dimensi yang berbeda untuk algoritma Lloyd yang dapat digunakan untuk mengurangi ketergantungan partisi	Knowledge-Based System	Science Direct
2	Liang Bai dkk. (2017) [16].	Fast density clustering strategies based on the K-Means algorithm	General	CFSDP Algorithm	Membutuhkan waktu komputasi lebih lama dan perhitungan jarak lebih banyak	Pattern Recognition	Science Direct
3	Agus Perdana Windarto. (2017) [17].	Implementation of Data Mining on Rice Imports by Major Country of Origin Using K-Means Clustering Method	Sosial	Euclidean Distance	Pengolahan data menggunakan Euclidean Distance memberikan bobot kriteria data yang lebih akurat dalam proses pengelompokan	International Journal of Artificial Intelligence Research	Google Scholar
4	Siwei Wang dkk. (2019) [18].	K-Means Clustering With Incomplete Data	General	Elbow Criterion	Secara dinamis dapat mengoptimalkan nilai pusat cluster dalam pengelompokan data yang tidak lengkap	IEEE Access	IEEE
5	Hailun Xie dkk. (2019) [19].	Improving K-Means clustering with enhanced Firefly Algorithm	General	Firefly Algorithm	Koefisien perhitungan jarak pada firefly algoritma memiliki strategi pengelempokan dalam satu	Applied Soft Computing	Science Direct

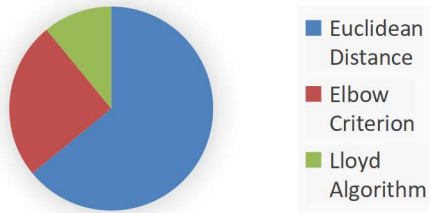
					dimensi dan multi dimensi sehingga skala pengelompokan lebih besar		
6	Yuda Irawan. (2019) [20].	Implementation Of Data Mining For Determining Majors Using K-Means Algorithm In Students Of SMA Negeri 1 Pangkalan Kerinci	Education	Euclidean Distance	Pengujian data menggunakan Euclidean Distance memberikan waktu komputasi lebih cepat dan lebih mudah dalam mengolah data	Journal of Applied Engineering and Technological Science	Google Scholar
7	Tao Li dkk. (2020) [21].	Normalization-Based Validity Index of Adaptive K-Means Clustering for Multi-Solution Application	General	NBVI Algorithm	Perhitungan jarak dengan metode NBVI berkinerja lebih baik dalam solusi parsial dan ganda perspektif.	IEEE Access	IEEE
8	Chang Xia dkk. (2020) [22]	Distributed K-Means clustering guaranteeing local differential privacy	General	Feature Algorithm	Algoritma feature menunjukkan bahwa eksperimen kinerja perhitungan jarak memiliki pengelompokan yang baik	Computers & Security	Science Direct
9	Junwen Chen dkk. (2020) [23].	Quantum-inspired ant lion optimized hybrid K-Means for cluster analysis and intrusion detection	General	Euclidean Distance	Pengolahan data menggunakan Euclidean Distance mampu meningkatkan kekurangan inisialisasi acak pada pusat data	Knowledge-Based System	Science Direct
10	Arvinder Kaur dkk. (2020) [24].	Hybridization of Chaos and Flower Pollination Algorithm over K-Means for data clustering	General	CFPA Algorithm	Memiliki rata-rata waktu lebih cepat untuk pengolahan data dan tingkat konvergensi yang cepat	Applied Soft Computing	Science Direct
11	Hong-Hao Zhao dkk. (2021) [25].	An Extended Regularized K-Means Clustering Approach for High-Dimensional Customer Segmentation With Correlated	General	Lloyd Algorithm	Mampu menghasilkan kesalahan pengelompokan yang lebih kecil dan dapat memilih variabel selama proses pengelompokan	IEEE Access	IEEE

		Variables					
12	Xi Wang dkk. (2021) [26].	An adaptive and opposite K-Means operation based memetic algorithm for data clustering	General	Elbow Criterion	Dapat meningkatkan efisiensi evolusi dan meningkatkan kapasitas eksplorasi, sehingga menghilangkan optima yang tidak perlu	Neurocomputing	Science Direct
13	Yongkai Fan dkk. (2021) [27].	PPMCK: Privacy-preserving multi-party computing for K-Means clustering	General	PPMCK Algorithm	Mampu meningkatkan efisiensi komputasi dan menjamin privasi data dalam cloud pada saat yang sama ketika interaksi, meskipun membutuhkan biaya lebih untuk enkripsi data	J. Parallel. Distrib. Compute	Science Direct
14	Atif Rizwan dkk. (2021) [28].	K-Means Clustering Algorithm for Groundwater Resource Planning in Point Cloud	Sosial	Euclidean Distance	Hasil menunjukkan bahwa kinerja algoritma Euclidean Distance dalam pengelompokan data yang diusulkan lebih baik dan kompetitif untuk mengimbangi pengelompokan yang diterapkan	IEEE Access	IEEE
15	Shudong Huang dkk. (2021) [29].	Robust deep K-Means: An effective and simple method for data clustering	General	RDKM Algorithm	Algoritma optimasi yang diusulkan memiliki kinerja pengelempokan yang kuat dan efektif, serta sangat cepat memusatkan data	Pattern Recognition	Science Direct

3.3 Perbandingan Metode Berpengaruh

Berdasarkan perbandingan dari literatur terpilih, gambar 4. menunjukkan bahwa sebagian besar metode atau teori perhitungan jarak pada

algoritma K-Means didominasi oleh Euclidean Distance, disusul dengan Elbow Criterion dan Lloyd Algorithm.



Gambar 4. Teori Pengukuran Jarak

Berdasarkan Tabel 5. dapat diketahui bahwa algoritma K-Means memiliki karakteristik sebagai berikut:

1. Dalam berbagai bidang studi, algoritma K-Means masih menjadi algoritma yang populer selama lima tahun terakhir yakni pada tahun 2017-2021, dibuktikan dengan banyaknya paper terkait yang terbit di berbagai publikasi jurnal bereputasi internasional.
2. Ditemukan teori perhitungan jarak yang baru pada algoritma K-Means.
3. Euclidean Distance menjadi teori algoritma perhitungan jarak yang masih banyak digunakan.
4. Hasil perbandingan metode menunjukkan bahwa Euclidean Distance memiliki keunggulan perhitungan jarak yang lebih baik.
5. Algoritma CFSDP memiliki kekurangan pada perhitungan jarak karena membutuhkan waktu komputasi lebih lama.
6. Publikasi jurnal pada *database* Science Direct terkait algoritma K-Means memiliki tingkat keunikan terkait teori perhitungan jarak yang baru dibandingkan dengan *database* IEEE dan Google Scholar.
7. Science Direct menjadi *database* yang mendominasi terbitnya paper terkait algoritma K-Means dibandingkan dengan *database* IEEE dan Google Scholar.

Proses suatu algoritma K-Means adalah sebagai berikut:

1. Menentukan banyaknya *cluster* yang akan dibuat serta ditetapkan sebagai awal dari pusat *cluster* k.
2. Menggunakan jarak Euclidean Distance yang kemudian dihitung dari setiap data ke awal pusat *cluster*

$$d(i, k) = \sqrt{\sum_i^m (C_{ij} - C_{kj})^2}$$

3. Kategorikan data ke dalam *cluster* dengan jarak yang terpendek dengan menggunakan

persamaan

$$\min \sum_k^i -aik - = \sqrt{\sum_i^m (C_{ij} - C_{kj})^2}$$

4. Hitung pusat *cluster* yang baru menggunakan persamaan

$$C_{kj} = \frac{\sum_k^i X_{ij}}{p} \quad (3)$$

Dengan : X^c *cluster* ke k p = banyaknya anggota *cluster* ke k

5. Ulangi langkah kedua sampai dengan keempat sehingga sudah tidak ada lagi data yang berpindah ke *cluster* yang lain.

4. KESIMPULAN DAN SARAN

Setelah melakukan penelitian, dapat disimpulkan bahwa: algoritma K-Means masih banyak digunakan dalam rentang waktu 2017-2021 terutama di bidang sosial. Berdasarkan pemilihan literatur publikasi jurnal berpengaruh, *Pattern Recognition*, *Knowledge-Based System*, *Applied Soft Computing*, dan *IEEE Access* dapat menjadi rujukan utama terkait algoritma K-Means, karena keempat publikasi jurnal tersebut merupakan jurnal dengan kualitas reputasi internasional diperkuat dengan ranking pada SJR. Metode atau teori perhitungan jarak pada algoritma K-Means yang sering digunakan yakni Euclidean Distance, Elbow Criterion, dan Lloyd Algorithm. Hasil perbandingan metode menunjukkan bahwa Euclidean Distance memiliki keunggulan perhitungan jarak yang lebih baik. Diharapkan penelitian ini dapat memotivasi peneliti selanjutnya untuk mengembangkan model penelitian yang lebih mendalam dan mengeksplorasi lebih lanjut terkait algoritma K-Means teknik *clustering* pada data mining karena dirasa masih sangat perlu dilakukan di masa depan.

DAFTAR PUSTAKA

- [1] S. Jambekar and Z. Saquib, "Prediction of Crop Production in India Using Data Mining Techniques," *2018 Fourth Int. Conf. Comput. Commun. Control Autom.*, pp. 1-5, 2018.
- [2] Y. Yin, L. Long, and X. Deng, "Dynamic Data Mining of Sensor Data," *IEEE Access*, vol. 8, pp. 41637-

- 41648, 2020, doi: 10.1109/ACCESS.2020.2976699.
- [3] G. Gustientiedina, M. H. Adiya, and Y. Desnelita, "Penerapan Algoritma K-Means Untuk Clustering Data Obat-Obatan," *J. Nas. Teknol. dan Sist. Inf.*, vol. 5, no. 1, pp. 17–24, 2019, doi: 10.25077/teknosi.v5i1.2019.17-24.
- [4] P. O. Olukanmi and B. Twala, "K-Means-sharp: Modified centroid update for outlier-robust K-Means clustering," *2017 Pattern Recognit. Assoc. South Africa Robot. Mechatronics Int. Conf. PRASA-RobMech 2017*, vol. 2018-Janua, pp. 14–19, 2017, doi: 10.1109/RoboMech.2017.8261116.
- [5] M. S. Yang and K. P. Sinaga, "A feature-reduction multi-view K-Means clustering algorithm," *IEEE Access*, vol. 7, pp. 114472–114486, 2019, doi: 10.1109/ACCESS.2019.2934179.
- [6] M. A. Mercioni and S. Holban, "Evaluating hierarchical and non-hierarchical grouping for develop a smart system," *2018 13th Int. Symp. Electron. Telecommun. ISETC 2018 - Conf. Proc.*, pp. 1–4, 2018, doi: 10.1109/ISETC.2018.8583997.
- [7] pallavi pallu, R. Suryawnashi, A. Dubey, and A. Abha Choubey, "A Systematic Review on K-Means Clustering Techniques Related papers A Novel Approach for Dat a Clust ering using Improved K-Means Algorit hm A Systematic Review on K-Means Clustering Techniques," *Int. J. Sci. Res. Eng. Technol.*, vol. 6, no. 6, 2017.
- [8] W. DIng, Y. Zhang, Y. Sun, and T. Qin, "An Improved SFLA-Kmeans Algorithm Based on Approximate Backbone and Its Application in Retinal Fundus Image," *IEEE Access*, vol. 9, pp. 72259–72268, 2021, doi: 10.1109/ACCESS.2021.3079119.
- [9] S. S. Yu, S. W. Chu, C. M. Wang, Y. K. Chan, and T. C. Chang, "Two improved K-Means algorithms," *Appl. Soft Comput. J.*, vol. 68, pp. 747–755, 2018, doi: 10.1016/j.asoc.2017.08.032.
- [10] C. Kamila, M. Adiyatma, G. R. Namang, R. Ramadhan, F. Syah, and D. Redaksi, "Pendidikan Teknik Informatika dan Komputer, Fakultas Teknik," *J. Intech*, vol. 2, no. 1, p. 1, 2021.
- [11] A. Ashabi, S. Bin Bin Sahibuddin, and M. Salkhordeh Salkhordeh Haghighi, "The systematic review of K-Means clustering algorithm," *ACM Int. Conf. Proceeding Ser.*, pp. 13–18, 2020, doi: 10.1145/3447654.3447657.
- [12] D. W. L. Pamungkas and S. Rochimah, "Pengujian Aplikasi Web - Tinjauan Pustaka Sistematis," *J. IPTEK*, vol. 23, no. 1, pp. 17–24, 2019, doi: 10.31284/j.ipitek.2019.v23i1.459.
- [13] R. S. Wahono, "A Systematic Literature Review of Software Defect Prediction: Research Trends, Datasets, Methods and Frameworks," *J. Softw. Eng.*, vol. 1, no. 1, pp. 1–16, 2015.
- [14] B. Kitchenham, O. Pearl Brereton, D. Budgen, M. Turner, J. Bailey, and S. Linkman, "Systematic literature reviews in software engineering - A systematic literature review," *Inf. Softw. Technol.*, vol. 51, no. 1, pp. 7–15, 2009, doi: 10.1016/j.infsof.2008.09.009.
- [15] M. Capó, A. Pérez, and J. A. Lozano, "An efficient approximation to the K-Means clustering for massive data," *Knowledge-Based Syst.*, vol. 117, pp. 56–69, 2017, doi: 10.1016/j.knosys.2016.06.031.
- [16] L. Bai, X. Cheng, J. Liang, H. Shen, and Y. Guo, "Fast density clustering strategies based on the K-Means algorithm," *Pattern Recognit.*, vol. 71, pp. 375–386, 2017, doi: 10.1016/j.patcog.2017.06.023.
- [17] A. P. Windarto, "Implementation of Data Mining on Rice Imports by Major Country of Origin Using Algorithm Using K-Means Clustering Method," *Int. J. Artif. Intell. Res.*, vol. 1, no. 2, p. 26, 2017, doi: 10.29099/ijair.v1i2.17.
- [18] S. Wang *et al.*, "K-Means Clustering With Incomplete Data," *IEEE Access*, vol. 7, pp. 69162–69171, 2019, doi: 10.1109/ACCESS.2019.2910287.
- [19] H. Xie *et al.*, "Improving K-Means clustering with enhanced Firefly Algorithms," *Appl. Soft Comput. J.*, vol. 84, p. 105763, 2019, doi: 10.1016/j.asoc.2019.105763.
- [20] Yuda Irawan, "Implementation Of Data Mining For Determining Majors Using K-Means Algorithm In Students Of SMA Negeri 1 Pangkalan Kerinci," *J. Appl. Eng. Technol. Sci.*, vol. 1, no. 1, pp. 17–29, 2019, doi:

- 10.37385/jaets.v1i1.18.
- [21] T. Li, Y. Ma, and T. Endoh, "Normalization-based validity index of adaptive K-Means clustering for multi-resolution application," *IEEE Access*, vol. 8, pp. 9403–9419, 2020, doi: 10.1109/ACCESS.2020.2964763.
- [22] C. Xia, J. Hua, W. Tong, and S. Zhong, "Distributed K-Means clustering guaranteeing local differential privacy," *Comput. Secur.*, vol. 90, 2020, doi: 10.1016/j.cose.2019.101699.
- [23] J. Chen, X. Qi, L. Chen, F. Chen, and G. Cheng, "Quantum-inspired ant lion optimized hybrid K-Means for cluster analysis and intrusion detection," *Knowledge-Based Syst.*, vol. 203, p. 106167, 2020, doi: 10.1016/j.knsys.2020.106167.
- [24] A. Kaur, S. K. Pal, and A. P. Singh, "Hybridization of Chaos and Flower Pollination Algorithm over K-Means for data clustering," *Appl. Soft Comput.*, vol. 97, no. xxxx, p. 105523, 2020, doi: 10.1016/j.asoc.2019.105523.
- [25] H. H. Zhao, X. C. Luo, R. Ma, and X. Lu, "An Extended Regularized K-Means Clustering Approach for High-Dimensional Customer Segmentation with Correlated Variables," *IEEE Access*, vol. 9, pp. 48405–48412, 2021, doi: 10.1109/ACCESS.2021.3067499.
- [26] X. Wang, Z. Wang, M. Sheng, Q. Li, and W. Sheng, "An adaptive and opposite K-Means operation based memetic algorithm for data clustering," *Neurocomputing*, vol. 437, pp. 131–142, 2021, doi: 10.1016/j.neucom.2021.01.056.
- [27] Y. Fan *et al.*, "PPMCK: Privacy-preserving multi-party computing for K-Means clustering," *J. Parallel Distrib. Comput.*, vol. 154, pp. 54–63, 2021, doi: 10.1016/j.jpdc.2021.03.009.
- [28] A. Rizwan, N. Iqbal, A. N. Khan, R. Ahmad, and D. H. Kim, "Toward Effective Pattern Recognition Based on Enhanced Weighted K-Mean Clustering Algorithm for Groundwater Resource Planning in Point Cloud," *IEEE Access*, vol. 9, pp. 130154–130169, 2021, doi: 10.1109/ACCESS.2021.3111112.
- [29] S. Huang, Z. Kang, Z. Xu, and Q. Liu, "Robust deep K-Means: An effective and simple method for data clustering," *Pattern Recognit.*, vol. 117, p. 107996, 2021, doi: 10.1016/j.patcog.2021.107996.