

Literature Review : Implementasi Algoritma K-Means Clustering Dalam Berbagai Sektor

Rivaldo Janter Tampubolon^{1*}, Thoty Ariandes², Muhammad Idlal Hafizd³, Najwan Hidayatullah⁴

^{1,2,3,4}Fakultas Ilmu Komputer, Program Studi Teknik Informatika, Universitas Pamulang, Tangerang Selatan Banten, Indonesia

Email: ^{1*}rivaldojnr@gmail.com, ²thotyariesandes.04@gmail.com, ³dalhafizd@gmail.com, ⁴najwanlf00@gmail.com

(* : coresponding author)

Abstrak – K-Means Clustering merupakan salah satu algoritma machine learning yang sangat penting dalam analisis data dan pengolahan informasi modern saat ini. Penelitian ini bertujuan untuk mengkaji penerapan algoritma *K-Means Clustering* melalui tinjauan sistematis terhadap jurnal ilmiah. Studi literatur ini mengeksplorasi implementasi K-Means dalam empat bidang berbeda: kesehatan masyarakat, pendidikan, pemasaran dan pertanian. Metode penelitian yang digunakan adalah analisis komparatif terhadap jurnal-jurnal yang telah dipublikasikan, dengan fokus pada metodologi, hasil, dan implikasi praktis dari penggunaan algoritma clustering. Hasil menunjukkan bahwa *K-Means* memiliki fleksibilitas tinggi dalam mengidentifikasi pola dan mendukung pengambilan keputusan lintas domain.

Kata Kunci: Implementasi *K-Means Clustering*; *Machine Learning*; Pengambilan Keputusan

Abstract – *K-Means Clustering* is a crucial machine learning algorithm that plays a significant role in modern data analysis and information processing. This research aims to examine the application of *K-Means Clustering* through a systematic review of scientific journals. The literature study explores the implementation of *K-Means* in four different fields: public health, education, marketing, and agriculture. The research methodology employed a comparative analysis of published journals, focusing on methodologies, results, and practical implications of clustering algorithm usage. The findings demonstrate that *K-Means* possesses high flexibility in identifying patterns and supporting cross-domain decision-making processes.

Keywords: *K-Means Clustering*; *Machine Learning*; *Decision-Making*

1. PENDAHULUAN

K-Means Clustering merupakan salah satu algoritma machine learning yang penting dalam analisis data. Penelitian ini bertujuan untuk mengkaji penerapan algoritma *K-Means Clustering* melalui tinjauan sistematis terhadap lima jurnal ilmiah. Studi literatur ini mengeksplorasi implementasi *K-Means* dalam berbagai bidang, termasuk kesehatan masyarakat, pendidikan, pemasaran, dan sektor lainnya. Metode penelitian yang digunakan adalah analisis komparatif terhadap jurnal-jurnal yang telah dipublikasikan, dengan fokus pada metodologi, hasil, dan implikasi praktis dari penggunaan algoritma *clustering*. Hasil menunjukkan bahwa *K-Means* memiliki fleksibilitas tinggi dalam mengidentifikasi pola dan mendukung pengambilan keputusan lintas domain.

Kelebihan utama algoritma *K-Means* adalah kemampuannya menangani variasi data yang kompleks. Metode ini dapat mengelompokkan data dengan cara memberikan penanda atau pusat kluster pada setiap kelompok data. Ini memungkinkan algoritma untuk menghasilkan pengelompokan yang akurat berdasarkan kedekatan karakteristik dan atribut data. Selain itu, *K-Means* memiliki keunggulan dalam memberikan perspektif baru dalam analisis data. Algoritma ini dapat menyediakan informasi tentang pola tersembunyi, membantu peneliti atau praktisi dalam membuat keputusan yang lebih terarah.

Namun, metode *K-Means* juga memiliki beberapa keterbatasan. Pertama, algoritma ini memerlukan penentuan jumlah kluster (*K*) di awal, yang membutuhkan pemahaman mendalam tentang struktur data. Proses ini membutuhkan ketelitian dan pengalaman dalam menganalisis dataset. Kedua, hasil clustering sangat bergantung pada inisialisasi pusat kluster awal, yang dapat menghasilkan hasil yang berbeda pada setiap running. Oleh karena itu, pemahaman yang komprehensif tentang karakteristik data dan teknik *preprocessing* sangat penting dalam

mengimplementasikan *K-Means*.

Penelitian ini bertujuan untuk menyelidiki implementasi algoritma *K-Means Clustering* dalam berbagai konteks, dengan harapan dapat memberikan gambaran komprehensif tentang fleksibilitas dan kemampuan metode ini. Dengan meningkatnya pemahaman tentang algoritma clustering, diharapkan penelitian ini dapat membantu peneliti dan praktisi dalam mengeksplorasi pola data yang kompleks. Para pengambil keputusan diharapkan mendapatkan bantuan dalam menganalisis data lintas bidang. Selain itu, harapannya adalah penelitian ini dapat menyumbangkan pengetahuan baru dalam pengembangan metode analisis data yang lebih efektif, dengan memanfaatkan kemampuan *K-Means* dalam mengungkap wawasan tersembunyi dari kumpulan data yang rumit.

2. METODE

Studi ini menggunakan pendekatan tinjauan pustaka untuk menyelidiki penerapan algoritma K-Means Clustering secara komprehensif. Metodologi penelitian difokuskan pada kajian literatur sistematis dengan menelaah publikasi ilmiah yang membahas penggunaan algoritma K-Means dalam konteks yang beragam. Proses pengumpulan referensi dilakukan melalui platform penelusuran akademik, dengan memanfaatkan mesin pencari seperti Google Scholar dan menggunakan variasi frasa kunci yang mencakup "K-Means Clustering", "teknik pengelompokan", "analitik data", serta "aplikasi K-Means". Tahap selanjutnya meliputi proses seleksi artikel berdasarkan relevansi dengan ruang lingkup penelitian. Melalui analisis mendalam terhadap sumber-sumber terpilih, penelitian ini bermaksud mengungkap fleksibilitas algoritma K-Means Clustering dalam mengidentifikasi pola dan melakukan agregasi data lintas berbagai ranah kajian.

3. ANALISA DAN PEMBAHASAN

3.1 Analisa

Tabel 1. Rangkuman Hasil Artikel

No.	Judul	Author / Tahun	Metode	Hasil
1.	Clustering of District or City in Central Java Based COVID-19 Case Using K-Means Clustering.	Ali Mahmudan. 2020	<i>K-Means</i>	<p>Penelitian ini menghasilkan pengelompokan kabupaten/kota di Provinsi Jawa Tengah berdasarkan kasus COVID-19 menggunakan metode K-Means Clustering. Dalam penelitian ini, terdapat tiga cluster yang terbentuk:</p> <ul style="list-style-type: none">Cluster 1: Terdiri dari 33 kabupaten/kota yang memiliki jumlah kasus aktif yang relatif rendah.Cluster 2: Terdiri dari 1 kota, yaitu Kota Semarang, yang menunjukkan jumlah kasus aktif yang sangat tinggi.Cluster 3: Terdiri dari 1 kabupaten, yaitu Kabupaten Jepara, yang juga memiliki jumlah kasus aktif yang signifikan. <p>Hasil analisis menunjukkan bahwa Cluster 2 dan Cluster 3 merupakan daerah yang perlu mendapatkan</p>

perhatian lebih dari pemerintah, karena keduanya memiliki jumlah kasus aktif dan kematian yang tinggi. Meskipun kondisi COVID-19 di Provinsi Jawa Tengah secara keseluruhan sudah relatif aman, penanganan yang maksimal tetap diperlukan untuk mencegah peningkatan jumlah kasus aktif dan kematian.

Sistem pengelompokan ini menjadi alat yang efektif untuk membantu pemerintah dalam mengambil keputusan yang tepat terkait penanganan COVID-19, serta memberikan informasi yang berguna bagi masyarakat dan tenaga medis dalam mengelola situasi pandemi.

2.	Implementasi Clustering Data Kasus Covid 19 Di Indonesia Menggunakan Algoritma KMeans	Nofita Sari, Hanny Hikmayanti, Handayani, Amril Mutoi Siregar. 2023	<i>K-Means</i>	Penelitian ini menghasilkan analisis pengelompokan data kasus Covid-19 di Indonesia menggunakan metode K-Means Clustering. Terdapat tiga cluster yang terbentuk berdasarkan jumlah kasus:
----	---	---	----------------	---

- Cluster 0: Terdiri dari 13 provinsi dengan jumlah kasus tergolong rendah, termasuk Nusa Tenggara Timur, Kepulauan Bangka Belitung, dan Sulawesi Tenggara.
- Cluster 1: Terdiri dari 16 provinsi dengan jumlah kasus paling tinggi, termasuk DKI Jakarta, Jawa Barat, dan Sumatera Utara.
- Cluster 2: Terdiri dari 5 provinsi dengan jumlah kasus sedang, termasuk Jawa Timur dan Kalimantan Timur.

Hasil analisis menunjukkan bahwa cluster 1 memiliki jumlah kasus tertinggi, sehingga dapat menjadi fokus untuk penanganan dan pengelolaan yang lebih intensif. Sementara itu, cluster 0 dan cluster 2 dapat difokuskan pada strategi pengelolaan yang lebih efisien. Dengan tingkat kecocokan sistem mencapai nilai optimal K: 3 dan *Sum of Square Error* (SSE) sebesar 228,913,736,548,657.56, metode *K-Means Clustering* terbukti efektif dalam pemetaan dan pengelompokan data kasus Covid-19, memberikan informasi

				berharga untuk pengambilan keputusan di sektor kesehatan.
3.	Analisis Data Argumen Tentang Penerapan Kebijakan Sistem Zonasi Pada Pendaftaran Sekolah Dengan Menggunakan K-Means Clustering	Ika Pusvitaningrum. 2020	<i>K-Means</i>	<p>Penelitian ini menghasilkan analisis tentang penerapan kebijakan sistem zonasi pada pendaftaran sekolah menggunakan metode <i>K-Means Clustering</i>. Terdapat lima cluster yang terbentuk berdasarkan pandangan masyarakat:</p> <ul style="list-style-type: none">• Cluster 1: Responden yang positif terhadap sistem zonasi.• Cluster 2: Responden yang kurang puas dengan penerapan zonasi.• Cluster 3: Responden yang memilih sekolah berdasarkan rekomendasi.• Cluster 4: Responden yang netral terhadap sistem zonasi.• Cluster 5: Responden yang merasa zonasi tidak berdampak positif. <p>Hasil analisis menunjukkan bahwa Cluster 2 dan Cluster 5 perlu perhatian lebih dari pemerintah, karena menunjukkan ketidakpuasan. Sistem pengelompokan ini membantu Dinas Pendidikan dalam pengambilan keputusan terkait penerapan sistem zonasi dan memberikan informasi berguna bagi masyarakat.</p>
4.	K-Means Clustering Method For Customer Segmentation Based On Potential Purchases	Baiq Nikum Yuliasih, Herman, Sunardi. 2024	<i>K-Means</i>	<p>Penelitian ini menghasilkan analisis segmentasi pelanggan menggunakan metode K-Means Clustering berdasarkan potensi pembelian di toko Fast Moving Consumer Goods (FMCG). Dalam penelitian ini, terdapat tiga cluster yang terbentuk berdasarkan karakteristik pelanggan:</p> <ul style="list-style-type: none">• Cluster 1 (k1): Terdiri dari 535 pelanggan dengan potensi pembelian rendah, umumnya adalah pria lajang, berusia rata-rata 37 tahun, dengan pendapatan tahunan sekitar \$97,107, dan tinggal di kota kecil.• Cluster 2 (k2): Terdiri dari 685 pelanggan dengan potensi pembelian tinggi, umumnya

adalah pria lajang, berusia rata-rata 39 tahun, dengan pendapatan tahunan sekitar \$143,402, dan tinggal di kota besar.

- Cluster 3 (k3): Terdiri dari 780 pelanggan dengan potensi pembelian sedang, umumnya adalah wanita menikah, berusia rata-rata 33 tahun, dengan pendapatan tahunan sekitar \$116,410, dan tinggal di kota menengah.

Hasil analisis menunjukkan bahwa cluster k2 memiliki potensi pembelian yang lebih tinggi, sehingga dapat menjadi target untuk penawaran produk eksklusif. Sementara itu, cluster k1 dan k3 dapat difokuskan pada strategi pemasaran yang lebih terjangkau dan disesuaikan dengan profil pelanggan mereka.

Sistem pengelompokan ini menjadi alat yang efektif untuk membantu toko FMCG dalam merancang strategi pemasaran yang lebih baik, meningkatkan manajemen inventaris, dan memberikan layanan yang lebih personal kepada pelanggan.

5.	Penerapan Metode K-Means Clustering untuk Pemetaan Pengelompokan Lahan Produksi Tandan Buah Segar.	Abdussalam Al Masykur, Siska Kurnia Gusti, Suwanto Sanjaya, Febi Yanto, Fadhilah Syafria. 2023	<i>K-Means</i>	Penelitian ini menghasilkan analisis pemetaan lahan produksi tandan buah segar (TBS) menggunakan metode K-Means Clustering di PT Surya Intisari Raya. Terdapat tiga cluster yang terbentuk berdasarkan produktivitas lahan: <ul style="list-style-type: none">• Cluster C1: Terdiri dari 96 blok lahan dengan produktivitas sedang.• Cluster C2: Terdiri dari 41 blok lahan dengan produktivitas rendah.• Cluster C3: Terdiri dari 79 blok lahan dengan produktivitas tinggi.
----	--	--	----------------	---

Hasil analisis menunjukkan bahwa cluster C3 memiliki produktivitas tertinggi, sehingga dapat menjadi fokus untuk peningkatan hasil produksi. Sementara itu, cluster C1 dan C2 dapat difokuskan pada strategi pengelolaan

yang lebih efisien. Dengan tingkat kecocokan sistem mencapai 99,54% dibandingkan dengan RapidMiner, metode K-Means Clustering terbukti efektif dalam pemetaan lahan produksi kelapa sawit, memberikan informasi berharga untuk pengelolaan dan pengambilan keputusan di sektor pertanian.

3.2 Pembahasan

Penelitian yang dilakukan oleh Ali Mahmudan mengenai pengelompokan kabupaten/kota di Provinsi Jawa Tengah berdasarkan kasus COVID-19 menunjukkan bahwa metode K-Means Clustering efektif dalam mengidentifikasi daerah yang memerlukan perhatian lebih dari pemerintah. Tiga cluster yang terbentuk memberikan gambaran jelas tentang distribusi kasus, di mana Cluster 2 dan Cluster 3, yang terdiri dari Kota Semarang dan Kabupaten Jepara, menunjukkan jumlah kasus aktif dan kematian yang tinggi. Temuan ini sejalan dengan penelitian sebelumnya yang menekankan pentingnya analisis clustering dalam pengambilan keputusan terkait penanganan pandemi, sehingga pemerintah dapat lebih fokus dalam alokasi sumber daya dan strategi penanganan yang tepat.

Selaras dengan itu, penelitian yang dilakukan oleh Nofita Sari, Hanny Hikmayanti Handayani, dan Amril Mutoi Siregar mengenai implemsentasi clustering data kasus COVID-19 di Indonesia menggunakan algoritma K-Means juga menunjukkan efektivitas metode ini dalam mengidentifikasi pola penyebaran COVID-19. Tiga cluster yang terbentuk memberikan wawasan tentang daerah dengan jumlah kasus tinggi, sedang, dan rendah, yang dapat membantu pemerintah dalam merumuskan strategi penanganan yang lebih tepat sasaran.

Di bidang pendidikan, penelitian oleh Ika Pusvitaningrum tentang penerapan kebijakan sistem zonasi pada pendaftaran sekolah mengungkapkan adanya variasi pandangan masyarakat yang signifikan terhadap kebijakan tersebut. Lima cluster yang terbentuk mencerminkan beragam perspektif orang tua, dari yang positif hingga skeptis. Penelitian ini menyoroti pentingnya memahami pandangan masyarakat untuk meningkatkan efektivitas kebijakan pendidikan. Aspek "Penerapan Kebijakan Sistem Zonasi" dan "Sistem Pendaftaran Sekolah" menjadi faktor kunci dalam pandangan masyarakat di Kecamatan Tayu, sementara di Kecamatan Margoyoso, "Pandangan Masyarakat Terhadap Label Sekolah" lebih dominan. Temuan ini menunjukkan bahwa kebijakan pendidikan harus mempertimbangkan konteks lokal dan melibatkan masyarakat dalam proses evaluasi untuk mencapai hasil yang lebih baik.

Dalam konteks bisnis, penelitian yang dilakukan oleh Baiq Nikum Yulisasih, Herman Herman, dan Sunardi Sunardi mengenai segmentasi pelanggan di toko FMCG menggunakan K-Means Clustering berhasil mengidentifikasi tiga cluster pelanggan berdasarkan potensi pembelian. Cluster dengan potensi pembelian tinggi menunjukkan karakteristik demografis yang berbeda dibandingkan dengan cluster lainnya, yang dapat dimanfaatkan untuk strategi pemasaran yang lebih terarah. Dengan memahami profil pelanggan, toko FMCG dapat merancang penawaran produk yang lebih sesuai dan meningkatkan kepuasan pelanggan. Hasil ini menunjukkan bahwa pendekatan berbasis data dalam segmentasi pasar dapat memberikan keuntungan kompetitif yang signifikan, yang sejalan dengan pentingnya analisis data dalam pengambilan keputusan di berbagai sektor.

Terakhir, penelitian yang dilakukan oleh Abdussalam Al Masykur dan rekan-rekannya juga menggunakan metode K-Means Clustering untuk memetakan lahan produksi tandan buah segar (TBS) di PT Surya Intisari Raya. Hasil penelitian menunjukkan bahwa metode ini dapat mengidentifikasi tiga cluster berdasarkan produktivitas lahan, dengan tingkat kecocokan sistem mencapai 99,54% dibandingkan dengan RapidMiner. Temuan ini menegaskan potensi K-Means Clustering dalam memberikan informasi berharga untuk pengelolaan dan pengambilan keputusan di sektor pertanian.

4. KESIMPULAN

Keseluruhan penelitian dari kelima artikel di atas menunjukkan bahwa penerapan metode K-Means Clustering memiliki potensi besar dalam analisis data di berbagai bidang. Penelitian pertama mengidentifikasi pengelompokan kabupaten/kota di Provinsi Jawa Tengah berdasarkan kasus COVID-19, yang memberikan informasi krusial bagi pemerintah dalam menentukan daerah yang memerlukan perhatian lebih. Penelitian kedua menyoroti penerapan kebijakan sistem zonasi pada pendaftaran sekolah, di mana analisis clustering memberikan wawasan mendalam tentang beragam pandangan masyarakat. Temuan ini menekankan pentingnya melibatkan masyarakat dalam evaluasi kebijakan pendidikan untuk mencapai hasil yang lebih baik..

Penelitian ketiga berfokus pada segmentasi pelanggan di toko FMCG, yang menunjukkan bahwa pemahaman karakteristik pelanggan dapat secara signifikan meningkatkan strategi pemasaran. Keunggulan metode K-Means Clustering dibandingkan dengan metode lainnya terletak pada kemampuannya untuk mengelompokkan data secara efektif, sehingga menghasilkan informasi yang berguna untuk pengambilan keputusan yang lebih tepat.

Meskipun hasil analisis menunjukkan potensi yang signifikan, masih diperlukan penelitian dan pengembangan lebih lanjut untuk meningkatkan akurasi dan efektivitas penerapan metode ini dalam konteks yang lebih luas. Oleh karena itu, eksplorasi lebih lanjut terhadap variabel yang mempengaruhi clustering dan penerapan teknik analisis yang lebih canggih dapat menjadi langkah penting untuk memaksimalkan manfaat dari metode ini.

Secara keseluruhan, penggunaan metode K-Means Clustering dalam kelima penelitian ini menunjukkan bahwa analisis data yang tepat dapat meningkatkan pengambilan keputusan di bidang kesehatan, pendidikan, bisnis dan pertanian. Dengan pemahaman yang lebih baik tentang dinamika yang ada, para pemangku kepentingan dapat merancang strategi yang lebih efektif dan responsif terhadap kebutuhan masyarakat, sehingga menciptakan dampak positif yang lebih besar dalam setiap sektor.

REFERENCES

- Al Masykur, A., Gusti, S.K., Sanjaya, S., Yanto, F., & Syafrina, F. (2023). Penerapan Metode K-Means Clustering untuk Pemetaan Pengelompokan Lahan Produksi Tandan Buah Segar. *Jurnal Informatika*.
- Baiq Nikum Yuliasih, Herman, H., & Sunardi, S. (2024). K-Means Clustering Method For Customer Segmentation Based On Potential Purchases. *Jurnal ELTIKOM : Jurnal Teknik Elektro, Teknologi Informasi Dan Komputer*, 8(1), 83–90. <https://doi.org/10.31961/eltikom.v8i1.1137>
- Mahmudan, A. (2020). Clustering of District or City in Central Java Based COVID-19 Case Using K-Means Clustering. *Jurnal Matematika, Statistika Dan Komputasi*, 17(1), 1–13. <https://doi.org/10.20956/jmsk.v17i1.10727>
- Pusvitaningrum, Ika. (2020). Analisis Data Argumen Tentang Penerapan Kebijakan Sistem Zonasi Pada Pendaftaran Sekolah Dengan Menggunakan K-Means Clustering. *Jurnal Buana Informatika*. 11. 1. 10.24002/jbi.v11i2.3575.
- Sari, Nofita & Hikmayanti, Hanny & Siregar, Amril. (2023). Implementasi Clustering Data Kasus Covid 19 Di Indonesia Menggunakan Algoritma K-Means. *Bianglala Informatika*. 11. 7-12. 10.31294/bi.v11i1.14762.