

IMPLEMENTASI K-MEANS UNTUK KLASTERISASI TINGKAT KEPARAHAN PENYAKIT PERIODONTAL BERDASARKAN DATA KLINIS

Lamria Pasaribu^{*1}, Khairunnisa Samosir², Yusra Fadhillah³

^{*1}Mahasiswa Ilmu Komputer, Fakultas Teknik, Universitas Graha Nusantara Padangsidempuan

^{2,3}Dosen Ilmu Komputer, Fakultas Teknik, Universitas Graha Nusantara Padangsidempuan

Email : ^{*1} lamriapsb04pasaribu@gmail.com

Abstract

Technological developments have had a significant impact in various fields, including health. In the field of dental health, one of the most common diseases is periodontal disease. Periodontal disease is a multifactorial disease that occurs in the supporting tissues of the teeth. The assessment of the severity of periodontal disease is often done subjectively. Therefore, a method is needed to help classify patient data based on the severity of periodontal disease in order to plan appropriate treatment based on the severity of the disease. This study uses the K-Means clustering algorithm. Clustering is performed based on gender, age, mean probing depth, mean clinical attachment loss, plaque index, and bleeding on probing. This study used the RapidMiner application and resulted in three levels of disease severity: mild, moderate, and severe. The evaluation results using the Davies-Bouldin Index showed a value of 0.688, indicating that the clusters formed had a good level of separation.

Keyword : *K-Means, Clustering, Periodontal disease*

Abstrak

Perkembangan teknologi telah membawa dampak signifikan dalam berbagai bidang, termasuk Kesehatan. Dalam bidang Kesehatan gigi, salah satu penyakit yang paling banyak diderita, ialah penyakit periodontal. Penyakit periodontal adalah penyakit multifaktorial yang terjadi pada jaringan penyangga gigi. Penilaian tingkat keparahan penyakit periodontal seringkali dilakukan secara subjektif. Sehingga dibutuhkan metode yang dapat membantu melakukan pengelompokan data pasien berdasarkan Tingkat keparahan penyakit periodontal agar dapat merencanakan perawatan yang sesuai berdasarkan Tingkat keparahan penyakitnya. Penelitian ini menggunakan algoritma K-Means clustering. Klasterisasi dilakukan berdasarkan jenis kelamin, usia, rata-rata kedalaman probing, rata-rata kehilangan pelekatan klinis, indeks plak, dan pendarahan pada probing. Penelitian ini menggunakan aplikasi rapidminer, dan menghasilkan 3 cluster tingkat keparahan, yaitu ringan, sedang, dan berat. Hasil evaluasi menggunakan Davies-Bouldin Index menunjukkan nilai sebesar 0.688, yang mengindikasikan bahwa klaster yang terbentuk memiliki tingkat pemisahan yang baik.

Kata Kunci: K-Means, Klastering, Penyakit Periodontal

PENDAHULUAN

Perkembangan teknologi telah membawa dampak signifikan dalam berbagai bidang, termasuk bidang Kesehatan. Teknologi menawarkan solusi inovatif untuk meningkatkan akurasi, dan kualitas pelayanan Kesehatan (Harahap et al., 2024). Misalnya sistem EMR (*Electronic Medical Record*) yang mempermudah pengelolaan data pasien (Firdonsyah & Nadia, 2024). *Telemedicine* yang memungkinkan konsultasi jarak jauh (Putri et al., 2024). Serta integrasi teknologi seperti *machine learning* yang dapat membantu dalam analisis data sehingga mempermudah diagnosis penyakit dan pengambilan keputusan klinis yang lebih akurat. (Akbar et al., 2025).

Kesehatan gigi dan mulut adalah masalah kesehatan masyarakat yang paling umum di dunia. Laporan Status Kesehatan Mulut (*World Health Organization*, 2022) menunjukkan bahwa hampir 3,5 miliar orang di seluruh dunia atau hampir setengah populasi dunia, mengalami gangguan pada kesehatan gigi dan mulut. Salah satu penyakit gigi dan mulut yang banyak diderita masyarakat global adalah penyakit periodontal atau gusi. Penyakit periodontal adalah penyakit multifaktorial yang terjadi pada jaringan penyangga gigi, melibatkan hubungan kompleks antara mikroorganisme oral dan imun tubuh. Kondisi ini menyebabkan kerusakan pada jaringan pendukung gigi, termasuk gusi, tulang alveolar, dan ligamen periodontal, yang pada akhirnya dapat menyebabkan kehilangan gigi (Kim et al., 2020). Penelitian epidemiologis menunjukkan adanya korelasi positif antara tingkat

keparahan penyakit periodontal dengan munculnya berbagai penyakit kronis, seperti penyakit kardiovaskular, diabetes, dan berdampak negatif pada kehamilan (Vollmer et al., 2022). Oleh karena itu deteksi awal keparahan penyakit periodontal merupakan langkah penting dalam menentukan rencana perawatan yang sesuai. Namun, penilaian tingkat keparahan penyakit periodontal seringkali dilakukan secara subjektif dan bergantung pada pengalaman tenaga medis, yang dapat memengaruhi hasil diagnosa dan perawatan pasien. Sehingga diperlukan metode yang lebih objektif dan efisien untuk menilai tingkat keparahan penyakit periodontal dan membantu tenaga medis dalam membuat keputusan perawatan yang lebih tepat.

Salah satu pendekatan yang dapat digunakan adalah analisis data menggunakan algoritma K-Means. K-means berfungsi memisahkan data ke dalam kluster yang berbeda berdasarkan kemiripan antar data (Azzahra & Amru Yasir, 2024). Penerapan algoritma K-Means telah banyak digunakan dalam bidang kesehatan, seperti Penelitian (Okta Jaya Harmaja, Hadirat Halawa, Wicarda Sandi Hulu, 2023) berjudul "Implementasi Algoritma K-Means Clustering Untuk Pengelompokan Penyakit Pasien Pada Puskesmas Pulo Brayan", telah berhasil menerapkan K-Means untuk mengelompokkan jenis penyakit pasien dengan menggunakan data pasien dan menghayang mencakup usia, jenis kelamin dan diagnosis penyakit. Sedikit berbeda dengan penelitian yang akan penulis

lakukan, penelitian ini berfokus pada implementasi algoritma K-means untuk mengelompokkan tingkat keparahan satu jenis penyakit yang lebih spesifik, yaitu

TINJAUAN PUSTAKA

1.

Penyakit Periodontal

Penyakit periodontal adalah infeksi bakteri kronis yang ditandai dengan kerusakan jaringan pendukung gigi seperti gingiva, ligament periodontal dan tulang alveolar (Azis, 2020). Periodontitis dapat muncul sebagai lanjutan dari gingivitis dan ditandai dengan peradangan kronis pada jaringan periodontal, yang mengganggu respon imun (Lee et al., 2025).

Penyebab utama penyakit periodontal adalah penumpukan plak. Plak terbentuk akibat interaksi antara sisa makanan dan bakteri yang ada di mulut. Plak ini kemudian berkembang menjadi karang gigi, yang menjadi tempat bagi bakteri untuk berkembang biak. Jika tidak dibersihkan, bakteri yang terdapat pada karang gigi dapat memproduksi racun yang menyebabkan peradangan dan iritasi pada area gusi, yang dikenal sebagai periodontitis. Jika tidak segera diobati, peradangan gusi dapat menciptakan celah antara jaringan gusi dan gigi. Hal ini memungkinkan bakteri untuk menginfeksi lebih dalam, berisiko merusak jaringan dan tulang di sekitar gusi. Akibatnya, gigi menjadi lebih rentan untuk tanggal dan dapat melemahkan sistem kekebalan tubuh (Siloam Hospitals, 2024).

penyakit periodontal dengan menggunakan data klinis pasien.

2. K-Means

K-Means merupakan salah satu algoritma klasterisasi yang cukup populer dan sederhana. Algoritma ini bertujuan untuk membagi sekumpulan data ke dalam K kelompok, dimana data dalam satu kelompok memiliki karakteristik yang lebih mirip satu sama lain dibandingkan dengan data di kelompok lainnya (Yang et al., 2023). Algoritma K-Means mengatur data ke dalam kluster berdasarkan kedekatannya dengan titik pusat kluster, yang disebut *centroid*.

3. Clustering

Clustering adalah proses pengelompokan data ke dalam cluster-cluster berdasarkan tingkat kemiripannya. Data dalam satu cluster harus memiliki kemiripan yang tinggi, sementara data antar cluster harus memiliki kemiripan yang rendah (Mauladi & Susilo, 2021). (Metisen & Sari, 2015) dalam (Saputra & Nataliani, 2021) menyebutkan Pengelompokan data diperlukan karena data mentah yang berukuran besar akan sulit untuk dianalisis dan dipahami secara langsung. Dalam konteks ini, clustering bertujuan untuk memperoleh pemahaman yang lebih baik tentang data serta mengevaluasi kualitasnya.

4. Data Mining

Data Mining juga dapat dibagi menjadi beberapa kategori atau teknik analisis yang berbeda, antara lain, karakteristik, diskriminasi, asosiasi, klasifikasi, clustering, trend, dan outlier (Pangestu & Ridwan, 2022). Data mining merupakan salah satu langkah kunci dalam upaya mengekstrak pengetahuan yang berguna dari database yang juga dikenal sebagai *Knowledge Discovery in Database* (KDD), merupakan serangkaian proses yang bertujuan untuk menemukan pola-pola penting atau menarik dari kumpulan data berukuran besar (Supoyo & Prasetyaningrum, 2022).

METODOLOGI PENELITIAN

Algoritma K-Means digunakan untuk mengelompokkan data klinis pasien penyakit periodontal berdasarkan tingkat keparahannya. Proses penelitian dilakukan melalui beberapa tahapan yang terstruktur guna memastikan data yang diolah bersih, akurat, serta menghasilkan klaster yang relevan dan dapat dimanfaatkan untuk analisis lanjutan.

Tahapan-tahapan yang diterapkan dalam penelitian ini meliputi:

1. Pengumpulan Data: data yang dikumpulkan berasal dari catatan medis pasien. mencakup informasi seperti kedalaman poket, kehilangan perlekatan klinis, indeks plak, dan pendarahan pada probing.
2. Pre-Proccesing Data: data klinis akan dibersihkan untuk menghilangkan *noise* atau inkonsistensi.

5. Rapidminer

RapidMiner merupakan perangkat lunak yang digunakan sebagai sarana pembelajaran dalam bidang data mining yang bersifat terbuka (*open source*). RapidMiner menyediakan sekitar 100 solusi pembelajaran untuk pengelompokan, klasifikasi, dan analisis regresi. Tak hanya itu, RapidMiner juga kompetibel dengan sekitar 22 format file, seperti .xls, .csv, dan sebagainya (Prasetyo et al., 2021). Serta dilengkapi sekitar 500 fitur yang mendukung proses penambahan informasi, dari input hingga representasi visual (Ananda et al., 2023).

3. Transformasi Data : data akan diubah ke format yang sesuai untuk input algoritma K-Means.

4. Implementasi K-Means : algoritma *K-Means* akan diimplementasikan menggunakan perangkat lunak rapidminer.

5. Hasil dan Visualisasi : hasil klasterisasi akan divisualisasikan menggunakan grafik atau diagram yang relevan untuk mempermudah interpretasi.

6. Evaluasi Hasil Clustering : untuk memastikan seberapa baik hasil clustering.

HASIL DAN PEMBAHASAN

4.1 Dataset

Data yang digunakan pada penelitian ini menggunakan data klinis pasien penyakit periodontal yang diperoleh dari praktik dokter gigi spesialis, drg. Erdi Efendi Nasution. Berikut penjelasan dataset yang digunakan.

Tabel 4.1.1 dataset

| NO. | ATRIBUT | DESKRIPSI |
|-----|---------------------------------------|---|
| 1. | Jenis Kelamin | Jenis kelamin pasien |
| 2. | Usia | Usia pasien saat pemeriksaan |
| 3. | Rata-rata kedalaman probing | kedalaman celah antar gusi dan gigi, nilai yang semakin tinggi menandakan kantong periodontal yang lebih dalam. diukur dengan satuan milimeter (mm) |
| 4. | Rata-rata kehilangan pelekatan klinis | menunjukkan seberapa banyak jaringan penyangga gigi yang telah hilang, diukur dengan satuan milimeter (mm) |
| 5. | Indeks plak | Persentase tingkat akumulasi plak pada gigi pasien |
| 6. | Pendarahan pada probing | Persentase terjadinya pendarahan pada gusi |

4.2 Transformasi Data

Data atribut yang akan diubah menjadi tipe numerik dan dilakukan pengkodean atau kategorikal nilai atribut pada data klinis pasien.

Tabel 4. Transformasi data pada atribut Jenis kelamin

| | |
|-----------|---|
| Perempuan | 1 |
| Laki-laki | 2 |

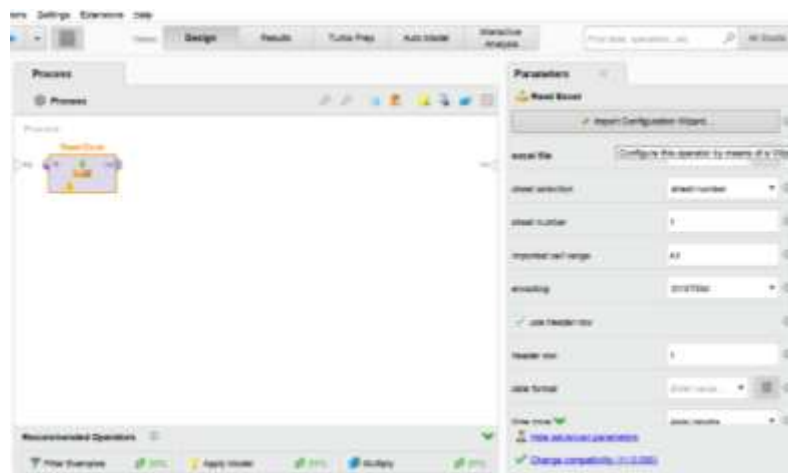
Tabel 4.1 Transformasi data pada atribut usia

| | | |
|---------------|--------------|------------|
| Kelompok Usia | Rentang Usia | Inialisasi |
|---------------|--------------|------------|

| | | |
|-------------------|-------|---|
| Masa Remaja Awal | 12-16 | 1 |
| Masa Remaja Akhir | 17-25 | |
| Masa dewasa awal | 26-35 | 2 |
| Masa dewasa akhir | 36-45 | |
| Masa lansia awal | 46-55 | 3 |
| Masa lansia akhir | 56-65 | |
| Manula | >65 | 4 |

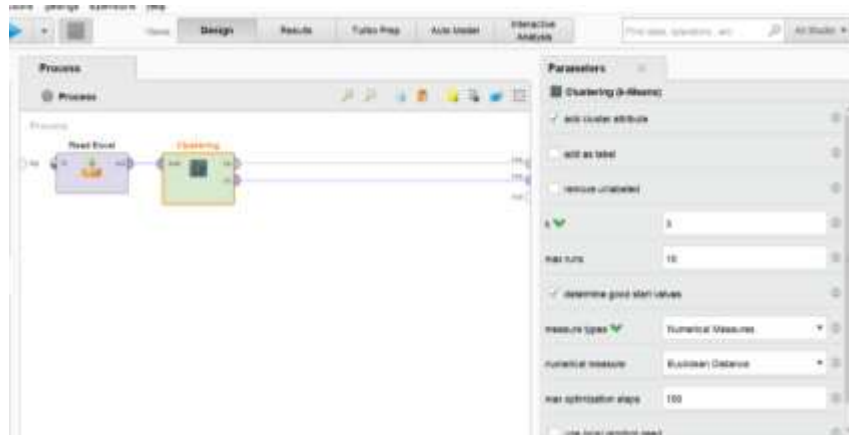
4.3 Implementasi K-Means

Tahapan awal dimulai dengan mengimport dataset yang telah disimpan sebelumnya dengan format excel (.xlsx). sehingga untuk membaca file ini digunakan operator *Read Excel*, kemudian *import* dataset menggunakan fitur *import configuration wizard* pada parameter *Read Excel*.



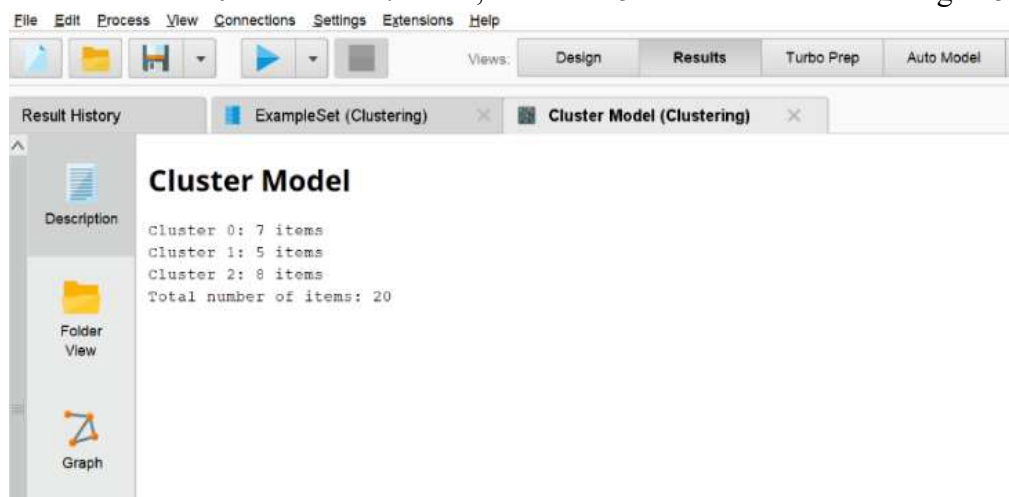
Gambar 4.3.1 operator read excel untuk mengimport dataset

Selanjutnya, tahap modeling menggunakan Algoritma K-Means. Data diproses menggunakan metode k-means dengan parameter $k = 3$ dan jenis pengukuran (numerical measure type). Hal ini dilakukan karena dataset sudah dalam bentuk numerik. Untuk menghitung jarak antara titik-titik cluster, digunakan Euclidean Distance sebagai standar.



Gambar 4.3.2 operator K-Means

Selanjutnya, data diproses menggunakan rapidminer dan menghasilkan 3 model cluster, sebagai berikut: cluster 0 terdiri dari 7 items, cluster 1 5 items dan cluster 2 dengan 8 items.



Gambar 4.3.3 Cluster Model

4.4 Interpretasi Hasil

| Attribute | cluster_0 | cluster_1 | cluster_2 |
|----------------------------|-----------|-----------|-----------|
| Jenis Kalam | 1.714 | 1.200 | 1.750 |
| Ura | 2 | 1.800 | 2.375 |
| Mean Probing Depth (mm) | 1.357 | 1.020 | 1.938 |
| Mean Attachment Level (mm) | -1.886 | -1.480 | -2.838 |
| Index Plus (%) | 57.714 | 39.600 | 74.350 |
| Bending On Probing (%) | 25.429 | 20 | 40.750 |

Gambar 4.4.1 Centroid Tabel

Centroid table menunjukkan nilai *centroid* (pusat) untuk masing-masing kluster berdasarkan atribut-atribut yang digunakan. Nilai *centroid* merepresentasikan rata-rata nilai dari seluruh data yang termasuk ke dalam masing-masing kluster.

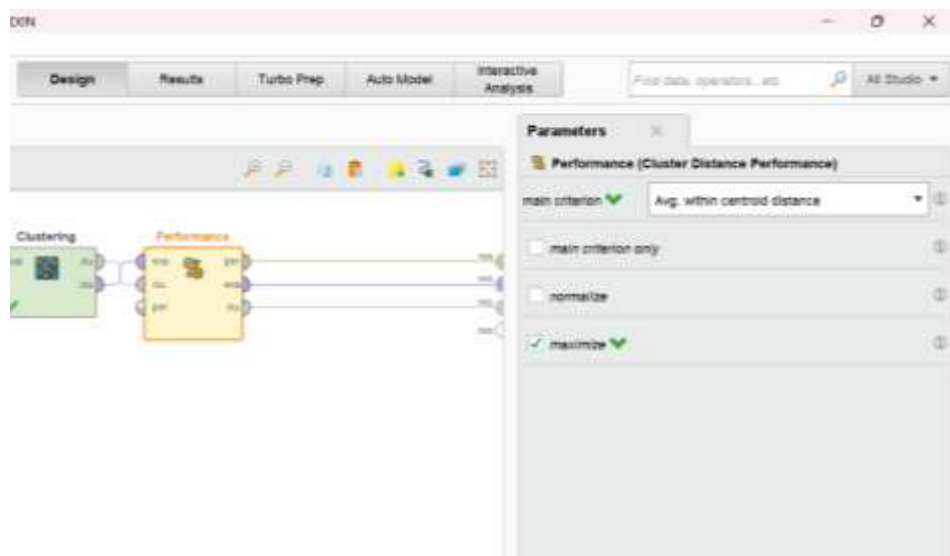
Setelah pembentukan kluster, didapat Kesimpulan sebagai berikut.

Tabel 4.4.1 Deskripsi data berdasarkan kluster

| Cluster | Deskripsi |
|---------|---|
| 0 | Anggota cluster dengan Tingkat keparahan Sedang |
| 1 | Anggota cluster dengan Tingkat keparahan Ringan |
| 2 | Anggota cluster dengan Tingkat keparahan Berat |

4.5 Evaluasi Hasil Clustering

Setelah proses clustering selesai, langkah selanjutnya adalah melakukan evaluasi untuk mengetahui kualitas hasil clustering atau efektivitas dari algoritma K-Means yang digunakan. Evaluasi ini dilakukan dengan memanfaatkan operator *Performance Vector* dalam RapidMiner.



Gambar 4.4.1 Operator Performance

Penilaian menggunakan metrik evaluasi Davies-Bouldin Index dan menghasilkan nilai numerik yang digunakan untuk menilai kualitas hasil klusterisasi. Nilai yang semakin mendekati nol menunjukkan bahwa kluster yang terbentuk semakin baik. Karena nilai Davies-Bouldin bersifat non-negatif (≥ 0), maka semakin kecil nilai yang diperoleh, semakin optimal kualitas klusterisasi tersebut.



Gambar 4.4.2 Nilai DBI

Diperoleh tingkat Performance akurasi Cluster sebesar 0.688.. Berdasarkan hasil evaluasi kinerja ini, dapat disimpulkan bahwa algoritma K-Means memiliki tingkat performa yang baik dalam melakukan pengelompokan data. pasien penyakit periodontal berdasarkan Tingkat keparahannya.

KESIMPULAN

1. Implementasi Algoritma K-means berhasil mengelompokkan data klinis pasien penyakit periodontal dengan parameter yang digunakan yaitu, jenis kelamin, usia, rata-rata kedalaman probing (*mean probing depth*), rata-rata kehilangan

pelekatan klinis (*mean clinical attachment level*), indeks plak (*index plak*), pendarahan pada probing (*bleeding on probing*) menjadi 3 kluster berdasarkan Tingkat keparahannya masing-masing.

2. Hasil evaluasi menggunakan metrix evaluasi *Davies-Bouldien Index* mendapatkan nilai DBI 0,688, hal ini menandakan bahwa model mampu membentuk kluster yang representatif, dengan penyebaran data yang cukup terpisah antar kluster.

SARAN

Berdasarkan kesimpulan hasil penelitian ini masih memiliki kelemahan yang

kedepannya perlu dikembangkan lagi. Beberapa saran yang diajukan :

1. Penelitian ini dapat dikembangkan dengan membandingkan algoritma K-Means dengan algoritma lainnya.
2. Penelitian ini dapat dikembangkan dengan mempertimbangkan penggunaan atribut yang lebih variatif atau menggunakan nilai yang lebih kompleks, seperti penggunaan nilai kedalaman probing per gigi serta atribut-atribut lain yang juga cukup krusial.
3. Penelitian ini berpotensi untuk dikembangkan menjadi sebuah sistem atau aplikasi yang dapat digunakan untuk mendukung proses pengambilan keputusan.

DAFTAR PUSTAKA

Akbar, I., Supriadi, F., & Junaedi, D. I. (2025). Pemanfaatan machine learning di bidang kesehatan. *JATI*

(*Jurnal Mahasiswa Teknik Informatika*), 9(1), 1744–1749.

Ananda, P. S., Sediono, E., & Sembiring, I. (2023). KMeans Clustering Menggunakan RapidMiner dalam Segmentasi Pelanggan dengan Evaluasi Davies Bouldin Index Untuk Menentukan Jumlah Cluster Paling Optimal. *Jurnal BATIRSI*, 6(2), 10.

Azis, N. A. A. (2020). Hubungan Penyakit Periodontal Dengan Penyakit Sistemik. *Kaos GL Dergisi*, 8(75), 147–154.
<https://doi.org/10.1016/j.jnc.2020.125798%0Ahttps://doi.org/10.1016/j.smr.2020.02.002%0Ahttp://www.ncbi.nlm.nih.gov/pubmed/810049%0Ahttp://doi.wiley.com/10.1002/anie.197505391%0Ahttp://www.sciencedirect.com/science/article/pii/B9780857090409500205%0Ahttp://>

Azzahra, L., & Amru Yasir. (2024). Metode K-Means Clustering Dalam Pengelompokan Penjualan Produk Frozen Food. *Jurnal Ilmu Komputer Dan Sistem Informasi*, 3(1), 1–10.
<https://doi.org/10.70340/jirsi.v3i1.88>

Firdonsyah, A., & Nadia, N. J. D. (2024). Sistem Rekam Medis Elektronik Berbasis Web Menggunakan Metode Waterfall. *Digital Transformation Technology*, 4(1), 636–643.
<https://doi.org/10.47709/digitech.v4i1.4482>

Harahap, F. A. A., Yulandari, M., Asshiddiqi, M. H., & Putri, H. (2024). Jurnal Kesehatan Unggul Gemilang. *Jurnal Kesehatan Unggul Gemilang*, 8(1), 7–15.

Kim, E. H., Kim, S., Kim, H. J., Jeong, H. O., Lee, J., Jang, J., Joo, J. Y., Shin, Y., Kang, J., Park, A. K., Lee, J. Y., & Lee, S. (2020). Prediction of Chronic Periodontitis Severity Using Machine Learning Models Based On

- Salivary Bacterial Copy Number. *Frontiers in Cellular and Infection Microbiology*, 10(November), 1–13. <https://doi.org/10.3389/fcimb.2020.571515>
- Lee, Y. H., Park, H. J., Jeong, S. J., Auh, Q. S., Jung, J., Lee, G. J., Shin, S., & Hong, J. Y. (2025). Oral microbiome profiles of gingivitis and periodontitis by next-generation sequencing among a group of hospital patients in Korea: A cross-sectional study. *Journal of Oral Biosciences*, 67(1), 100591. <https://doi.org/10.1016/j.job.2024.100591>
- Mauladi, K. F., & Susilo, P. H. (2021). Klusterisasi Virus Covid-19 Di Wilayah Kabupaten Lamongan Dengan Metode K-Means Clustering. *JUPI (Jurnal Ilmiah Penelitian Dan Pembelajaran Informatika)*, 6(2), 325–335. <https://doi.org/10.29100/jipi.v6i2.1999>
- Metisen, B. M., & Sari, H. L. (2015). Analisis clustering menggunakan metode K-Means dalam pengelompokan penjualan produk pada Swalayan Fadhila. *Jurnal Media Infotama*, 11(2), 110–118.
- Okta Jaya Harmaja, Hadirat Halawa, Wicarda Sandi Hulu, S. L. (2023). Implementasi Algoritma K-Means Clustering Untuk Pengelompokan Penyakit Pasien Pada Puskesmas Pulo Brayan. *Sains Dan Teknologi*, 5(1), 150–157.
- Pangestu, A., & Ridwan, T. (2022). Penerapan Data Mining Menggunakan Algoritma K-Means Pengelompokan Pelanggan Berdasarkan Kubikasi Air Terjual Menggunakan Weka. *JUST IT: Jurnal Sistem Informasi, Teknologi Informasi Dan Komputer*, 11(3), 67–71. <https://jurnal.umj.ac.id/index.php/just-it/article/view/11591>
- Prasetyo, V. R., Lazuardi, H., Mulyono, A. A., & Lauw, C. (2021). Penerapan Aplikasi RapidMiner Untuk Prediksi Nilai Tukar Rupiah Terhadap US Dollar Dengan Metode Linear Regression. *Jurnal Nasional Teknologi Dan Sistem Informasi*, 7(1), 8–17. <https://doi.org/10.25077/teknosi.v7i1.2021.8-17>
- Putri, J., Prayuti, Y., Triyana, Y., Farhan, M. I., Lany, A., & Fahrudin, A. (2024). Prespektif Hukum Kesehatan Dalam Penyelenggaraan Layanan Kesehatan Berbasis Digital Kedokteran Jarak Jauh (Telemedicine). *Jurnal Rectum: Tinjauan Yuridis Penanganan Tindak Pidana*, 6(2), 255–265.
- Saputra, E. A., & Nataliani, Y. (2021). Analisis Pengelompokan Data Nilai Siswa untuk Menentukan Siswa Berprestasi Menggunakan Metode Clustering K-Means. *Journal of Information Systems and Informatics*, 3(3), 424–439. <https://doi.org/10.51519/journalisi.v3i3.164>
- Supoyo, A., & Prasetyaningrum, P. T. (2022). Analisis Data Mining Untuk Memprediksi Lama Perawatan Pasien Covid-19 Di DIY. *Bianglala Informatika*, 10(1), 21–29. <https://doi.org/10.31294/bi.v10i1.11890>
- Vollmer, A., Vollmer, M., Lang, G., Straub, A., Shavlokhova, V., Kübler, A., Gubik, S., Brands, R., Hartmann, S., & Saravi, B. (2022). Associations between Periodontitis and COPD: An Artificial Intelligence-Based Analysis of NHANES III. *Journal of Clinical*

- Medicine*, 11(23).
<https://doi.org/10.3390/jcm11237210>
- World Health Organization. (2022). Global oral health status report. In *Who*, (Vol. 57, Issue 2).
- Yang, M., Huang, L., & Tang, C. (2023). K-Means Clustering with Local Distance Privacy. *Big Data Mining and Analytics*, 6(4), 433–442. <https://doi.org/10.26599/BDMA.2022.9020050>