



E-ISSN 3032-601X & P-ISSN 3032-7105

Vol. 3, No. 1, 2026

MISTER

**Journal of Multidisciplinary Inquiry in Science,
Technology and Educational Research**

**Jurnal Penelitian Multidisiplin dalam Ilmu
Pengetahuan, Teknologi dan Pendidikan**

**UNIVERSITAS SERAMBI MEKKAH
KOTA BANDA ACEH**

mister@serambimekkah.ac.id

Journal of Multidisciplinary Inquiry in Science Technology
and Educational Research

Journal of MISTER

Vol. 3, No. 1, 2026

Pages: 1025-1032

Analisis Pola Risiko Stroke pada Data Pasien Menggunakan Model K-Means Clustering Berbasis RapidMiner

Arjuna Septian Habibullah, Hasbi Firmansyah, Ali Sofyan, Wahyu
Asriyani

Universitas Pancasakti Tegal

Article in Journal of MISTER

Available at : <https://jurnal.serambimekkah.ac.id/index.php/mister/index>

DOI : <https://doi.org/10.32672/mister.v3i1.4026>

How to Cite this Article

APA : Habibullah, A. S., Firmansyah, H., Sofyan, A. ., & Asriyani, W. (2025). Analisis Pola Risiko Stroke pada Data Pasien Menggunakan Model K-Means Clustering Berbasis RapidMiner. Journal of Multidisciplinary Inquiry in Science, Technology and Educational Research, 3(1), 1025 - 1032. Retrieved from <https://jurnal.serambimekkah.ac.id/index.php/mister/article/view/4026>

Others Visit : <https://jurnal.serambimekkah.ac.id/index.php/mister/index>

MISTER: Journal of Multidisciplinary Inquiry in Science, Technology and Educational Research is a scholarly journal dedicated to the exploration and dissemination of innovative ideas, trends and research on the various topics include, but not limited to functional areas of Science, Technology, Education, Humanities, Economy, Art, Health and Medicine, Environment and Sustainability or Law and Ethics.

MISTER: Journal of Multidisciplinary Inquiry in Science, Technology and Educational Research is an open-access journal, and users are permitted to read, download, copy, search, or link to the full text of articles or use them for other lawful purposes. Articles on Journal of MISTER have been previewed and authenticated by the Authors before sending for publication. The Journal, Chief Editor, and the editorial board are not entitled or liable to either justify or responsible for inaccurate and misleading data if any. It is the sole responsibility of the Author concerned.



ISSN 3032-7105

ISSN 3032-601X



Analisis Pola Risiko Stroke pada Data Pasien Menggunakan Model K-Means Clustering Berbasis RapidMiner

Arjuna Septian Habibullah¹, Hasbi Firmansyah², Ali Sofyan³, Wahyu Asriyani⁴

Program Studi Informatika, Fakultas Teknik dan Ilmu Komputer, Universitas Pancasakti Tegal, Tegal, Indonesia^{1,2,3}

Program Studi Pendidikan Bahasa dan Sastra Indonesia, Fakultas Keguruan dan Ilmu Pendidikan, Universitas Pancasakti Tegal, Tegal, Indonesia⁴

*Email: arjunaseptian4@mail.com, hasbifirmansyah@upstegal.ac.id, alisofyan.bbs@gmail.com, asriyani1409@gmail.com.

Diterima: 10-12-2025

| Disetujui: 20-12-2025

| Diterbitkan: 22-12-2025

Abstract

*This study aims to identify patterns of stroke risk among patients by applying the K-Means clustering method using the RapidMiner platform. The stroke dataset was loaded from an Excel file and prepared through a preprocessing stage, including normalization of numerical attributes and conversion of categorical attributes into numerical form to ensure accurate distance calculations in the K-Means algorithm. The K-Means algorithm was then applied to form several clusters based on similarities in patients' clinical and demographic data. The quality of the clustering results was evaluated using the **Davies–Bouldin Index** to measure cluster separation and intra-cluster similarity. The results indicate that the K-Means method is capable of producing a more structured representation of stroke risk patterns and can support data-driven decision-making for stroke prevention and management.*

Keywords: K-Means Clustering; Stroke; RapidMiner; Data Mining; Davies–Bouldin Index.

Abstrak

Penelitian ini bertujuan untuk mengidentifikasi pola risiko stroke pada pasien menggunakan metode klastering K-Means dengan bantuan platform RapidMiner. Dataset stroke dimuat dari file Excel dan dipersiapkan melalui tahap preprocessing yang meliputi normalisasi atribut numerik serta konversi atribut kategorikal ke bentuk numerik agar perhitungan jarak K-Means dapat dilakukan secara akurat. Selanjutnya, algoritma K-Means diterapkan untuk membentuk beberapa cluster berdasarkan kesamaan data klinis dan demografis pasien. Kualitas hasil klastering dievaluasi menggunakan **Davies–Bouldin Index** untuk menilai tingkat pemisahan antar cluster dan keseragaman data dalam satu cluster. Hasil penelitian menunjukkan bahwa metode K-Means mampu menghasilkan pola risiko stroke yang lebih terstruktur dan dapat mendukung pengambilan keputusan pencegahan serta penanganan stroke berbasis data.

Katakunci: Stroke; K-Means; Klastering; RapidMiner; Davies–Bouldin Index; Data Mining.

PENDAHULUAN

Stroke merupakan salah satu penyebab utama kematian dan kecacatan secara global yang menuntut strategi deteksi dini berbasis pola risiko pasien agar penanganannya lebih efektif dan terarah (Jiang et al., n.d.). Kondisi ini menjadi tantangan serius bagi sistem kesehatan karena angka kejadian stroke terus meningkat seiring bertambahnya usia populasi dan perubahan gaya hidup. Di banyak negara, stroke tidak hanya menimbulkan beban biaya pengobatan yang besar, tetapi juga berdampak pada kualitas hidup pasien serta keluarga yang merawat.

Identifikasi pola risiko stroke memerlukan analisis multidimensi, karena kondisi tersebut dipengaruhi oleh faktor demografis, klinis, dan gaya hidup, sehingga metode analisis data berbasis machine learning semakin banyak digunakan dalam penelitian medis (Care et al., 2024). Faktor-faktor seperti usia, jenis kelamin, tekanan darah, kadar gula darah, hingga kebiasaan merokok memiliki keterkaitan yang beragam terhadap risiko terjadinya stroke, sehingga analisis konvensional seringkali tidak cukup untuk menangkap pola kompleks tersebut. Machine learning memungkinkan eksplorasi hubungan tersembunyi antara variabel dan dapat menemukan pola yang sulit diidentifikasi melalui pendekatan statistik tradisional. Salah satu pendekatan data mining yang umum digunakan adalah teknik *unsupervised learning* seperti clustering, yang memungkinkan peneliti menemukan struktur laten dan pengelompokan alami pada data tanpa memerlukan label hasil diagnosis (Kim et al., 2022). Pendekatan ini cocok untuk dataset kesehatan yang sering kali tidak memiliki label klasifikasi, karena diagnosis medis tidak selalu terdokumentasi secara lengkap dalam sistem data rumah sakit. Clustering juga membantu memetakan sub-kelompok pasien dalam populasi sehingga intervensi medis dapat dilakukan berdasarkan kelompok risiko tertentu.

Algoritma K-Means dikenal sebagai metode clustering yang sederhana, cepat, dan mampu memberikan centroid yang mudah ditafsirkan untuk mendukung pengambilan keputusan klinis (Ceskoutsé et al., 2025). Selain berfungsi untuk membagi data ke dalam beberapa kelompok homogen, K-Means juga memberikan representasi matematis dari pusat kelompok yang dapat dianalisis oleh tenaga medis untuk menilai karakteristik pasien. Dengan efisiensi perhitungan, algoritma ini sangat cocok digunakan pada dataset menengah, termasuk data kesehatan berbasis Excel yang banyak digunakan dalam penelitian akademik.

Studi terbaru menunjukkan bahwa penerapan K-Means dapat mengidentifikasi fenotipe kelompok pasien stroke berdasarkan kemiripan karakteristik, sehingga dapat mendukung strategi intervensi klinis yang lebih personal (Mao, 2025). Hal ini penting karena penanganan stroke bukan hanya tentang mengobati gejala akut, tetapi juga mengelola faktor risiko jangka panjang. Dengan mengetahui kelompok pasien yang memiliki kemungkinan risiko tinggi, tenaga medis dapat memberikan edukasi pencegahan dan perlakuan lebih intensif.

Selain itu, penggunaan model clustering terbukti efektif dalam memetakan pola multimorbiditas atau faktor risiko yang saling terkait, terutama pada pasien stroke dengan kondisi komorbid seperti hipertensi, diabetes mellitus, dan penyakit jantung (Guo et al., 2025). Multimorbiditas sangat umum terjadi pada pasien stroke dan memengaruhi jalur pengobatan serta proses rehabilitasi setelah serangan stroke. Dengan pengelompokan yang tepat, manajemen klinis dapat mengidentifikasi jenis komorbiditas dominan dalam populasi tertentu.

Hasil clustering membantu tenaga medis memahami profil pasien yang berbeda dari populasi umum sehingga dapat memengaruhi prioritas penanganan medis dan edukasi kesehatan (Id et al., 2023). Misalnya, pola yang ditemukan dapat digunakan untuk membuat program pencegahan dan intervensi yang

lebih tepat sasaran sesuai kelompok risiko masing-masing. Dengan analisis pola, tenaga medis dapat memfokuskan program rehabilitasi pada pasien dengan karakteristik klinis tertentu.

Penelitian terkait juga menunjukkan bahwa K-Means dapat diimplementasikan secara efektif menggunakan alat data mining visual seperti RapidMiner yang memberikan alur kerja terstruktur mulai dari preprocessing, normalisasi, pemilihan jumlah cluster (k), hingga visualisasi hasil (Rumah et al., 2024). RapidMiner menyediakan fitur grafik dan visualisasi yang memudahkan interpretasi hasil clustering tanpa perlu menulis kode pemrograman. Ini memungkinkan peneliti kesehatan dengan latar belakang non-teknis untuk melakukan analisis data secara mandiri.

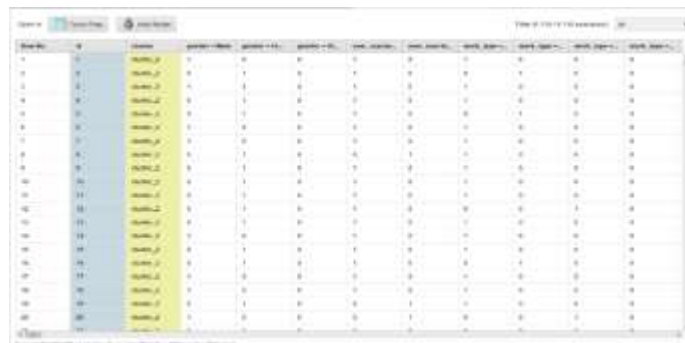
RapidMiner memungkinkan analisis data secara cepat tanpa memerlukan kemampuan pemrograman yang intensif, sehingga cocok digunakan dalam riset kesehatan berbasis data rumah sakit atau dataset Excel lokal (K-means, 2024). Selain itu, integrasi dengan berbagai format data seperti CSV dan Excel menjadikan RapidMiner fleksibel untuk penelitian dengan sumber data terbatas. Ini membuka peluang bagi institusi kesehatan untuk memanfaatkan data yang mereka miliki untuk meningkatkan kualitas layanan medis.

Berdasarkan urgensi tersebut, penelitian ini bertujuan untuk menganalisis pola risiko stroke dengan menerapkan model K-Means clustering berbasis RapidMiner pada dataset pasien, sehingga hasilnya diharapkan memberikan insight mengenai pengelompokan risiko stroke dan mendukung strategi pencegahan medis berbasis data (Wala & Umar, 2024). Temuan penelitian ini diharapkan dapat memberikan kontribusi pada metode deteksi dini berbasis data dan membantu institusi kesehatan dalam memahami distribusi risiko stroke pada populasi pasien tertentu.

METODE PENELITIAN

Penelitian ini bertujuan untuk mengidentifikasi dan menganalisis pola risiko stroke pada pasien dengan menerapkan metode klastering K-Means menggunakan perangkat lunak RapidMiner. Metode penelitian disusun secara bertahap dan sistematis agar setiap proses analisis dapat dilakukan secara terkontrol serta menghasilkan pola klaster yang valid dan dapat dipertanggungjawabkan secara ilmiah.

1. Pengumpulan Dataset

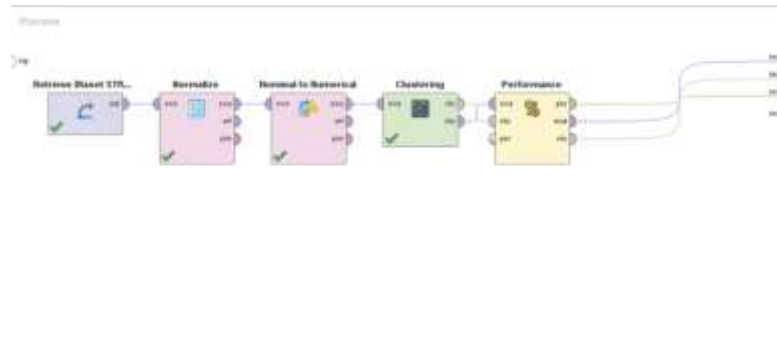


ID	Nama	Jenis Kelamin	Usia	Tekanan Darah	Kolesterol	Gula Darah	Riwayat Penyakit Jantung	Riwayat Penyakit Stroke	Riwayat Penyakit Lain
1	Andi, A	P	45	120/80	180	100	Ya	Tidak	Tidak
2	Budi, B	P	55	130/90	200	120	Tidak	Tidak	Tidak
3	Cici, C	P	65	140/100	220	140	Tidak	Tidak	Tidak
4	Dani, D	P	75	150/110	240	160	Tidak	Tidak	Tidak
5	Evi, E	P	85	160/120	260	180	Tidak	Tidak	Tidak
6	Fani, F	P	95	170/130	280	200	Tidak	Tidak	Tidak
7	Gani, G	P	105	180/140	300	220	Tidak	Tidak	Tidak
8	Hani, H	P	115	190/150	320	240	Tidak	Tidak	Tidak
9	Iani, I	P	125	200/160	340	260	Tidak	Tidak	Tidak
10	Jani, J	P	135	210/170	360	280	Tidak	Tidak	Tidak
11	Kani, K	P	145	220/180	380	300	Tidak	Tidak	Tidak
12	Lani, L	P	155	230/190	400	320	Tidak	Tidak	Tidak
13	Mani, M	P	165	240/200	420	340	Tidak	Tidak	Tidak
14	Nani, N	P	175	250/210	440	360	Tidak	Tidak	Tidak
15	Oani, O	P	185	260/220	460	380	Tidak	Tidak	Tidak
16	Pani, P	P	195	270/230	480	400	Tidak	Tidak	Tidak
17	Qani, Q	P	205	280/240	500	420	Tidak	Tidak	Tidak
18	Rani, R	P	215	290/250	520	440	Tidak	Tidak	Tidak
19	Sani, S	P	225	300/260	540	460	Tidak	Tidak	Tidak
20	Tani, T	P	235	310/270	560	480	Tidak	Tidak	Tidak
21	Uani, U	P	245	320/280	580	500	Tidak	Tidak	Tidak
22	Vani, V	P	255	330/290	600	520	Tidak	Tidak	Tidak
23	Wani, W	P	265	340/300	620	540	Tidak	Tidak	Tidak
24	Xani, X	P	275	350/310	640	560	Tidak	Tidak	Tidak
25	Yani, Y	P	285	360/320	660	580	Tidak	Tidak	Tidak
26	Zani, Z	P	295	370/330	680	600	Tidak	Tidak	Tidak
27	Aani, A	P	305	380/340	700	620	Tidak	Tidak	Tidak
28	Bani, B	P	315	390/350	720	640	Tidak	Tidak	Tidak
29	Cani, C	P	325	400/360	740	660	Tidak	Tidak	Tidak
30	Dani, D	P	335	410/370	760	680	Tidak	Tidak	Tidak
31	Eani, E	P	345	420/380	780	700	Tidak	Tidak	Tidak
32	Fani, F	P	355	430/390	800	720	Tidak	Tidak	Tidak
33	Gani, G	P	365	440/400	820	740	Tidak	Tidak	Tidak
34	Hani, H	P	375	450/410	840	760	Tidak	Tidak	Tidak
35	Iani, I	P	385	460/420	860	780	Tidak	Tidak	Tidak
36	Jani, J	P	395	470/430	880	800	Tidak	Tidak	Tidak
37	Kani, K	P	405	480/440	900	820	Tidak	Tidak	Tidak
38	Lani, L	P	415	490/450	920	840	Tidak	Tidak	Tidak
39	Mani, M	P	425	500/460	940	860	Tidak	Tidak	Tidak
40	Nani, N	P	435	510/470	960	880	Tidak	Tidak	Tidak
41	Oani, O	P	445	520/480	980	900	Tidak	Tidak	Tidak
42	Pani, P	P	455	530/490	1000	920	Tidak	Tidak	Tidak
43	Qani, Q	P	465	540/500	1020	940	Tidak	Tidak	Tidak
44	Rani, R	P	475	550/510	1040	960	Tidak	Tidak	Tidak
45	Sani, S	P	485	560/520	1060	980	Tidak	Tidak	Tidak
46	Tani, T	P	495	570/530	1080	1000	Tidak	Tidak	Tidak
47	Uani, U	P	505	580/540	1100	1020	Tidak	Tidak	Tidak
48	Vani, V	P	515	590/550	1120	1040	Tidak	Tidak	Tidak
49	Wani, W	P	525	600/560	1140	1060	Tidak	Tidak	Tidak
50	Xani, X	P	535	610/570	1160	1080	Tidak	Tidak	Tidak
51	Yani, Y	P	545	620/580	1180	1100	Tidak	Tidak	Tidak
52	Zani, Z	P	555	630/590	1200	1120	Tidak	Tidak	Tidak
53	Aani, A	P	565	640/600	1220	1140	Tidak	Tidak	Tidak
54	Bani, B	P	575	650/610	1240	1160	Tidak	Tidak	Tidak
55	Cani, C	P	585	660/620	1260	1180	Tidak	Tidak	Tidak
56	Dani, D	P	595	670/630	1280	1200	Tidak	Tidak	Tidak
57	Eani, E	P	605	680/640	1300	1220	Tidak	Tidak	Tidak
58	Fani, F	P	615	690/650	1320	1240	Tidak	Tidak	Tidak
59	Gani, G	P	625	700/660	1340	1260	Tidak	Tidak	Tidak
60	Hani, H	P	635	710/670	1360	1280	Tidak	Tidak	Tidak
61	Iani, I	P	645	720/680	1380	1300	Tidak	Tidak	Tidak
62	Jani, J	P	655	730/690	1400	1320	Tidak	Tidak	Tidak
63	Kani, K	P	665	740/700	1420	1340	Tidak	Tidak	Tidak
64	Lani, L	P	675	750/710	1440	1360	Tidak	Tidak	Tidak
65	Mani, M	P	685	760/720	1460	1380	Tidak	Tidak	Tidak
66	Nani, N	P	695	770/730	1480	1400	Tidak	Tidak	Tidak
67	Oani, O	P	705	780/740	1500	1420	Tidak	Tidak	Tidak
68	Pani, P	P	715	790/750	1520	1440	Tidak	Tidak	Tidak
69	Qani, Q	P	725	800/760	1540	1460	Tidak	Tidak	Tidak
70	Rani, R	P	735	810/770	1560	1480	Tidak	Tidak	Tidak
71	Sani, S	P	745	820/780	1580	1500	Tidak	Tidak	Tidak
72	Tani, T	P	755	830/790	1600	1520	Tidak	Tidak	Tidak
73	Uani, U	P	765	840/800	1620	1540	Tidak	Tidak	Tidak
74	Vani, V	P	775	850/810	1640	1560	Tidak	Tidak	Tidak
75	Wani, W	P	785	860/820	1660	1580	Tidak	Tidak	Tidak
76	Xani, X	P	795	870/830	1680	1600	Tidak	Tidak	Tidak
77	Yani, Y	P	805	880/840	1700	1620	Tidak	Tidak	Tidak
78	Zani, Z	P	815	890/850	1720	1640	Tidak	Tidak	Tidak
79	Aani, A	P	825	900/860	1740	1660	Tidak	Tidak	Tidak
80	Bani, B	P	835	910/870	1760	1680	Tidak	Tidak	Tidak
81	Cani, C	P	845	920/880	1780	1700	Tidak	Tidak	Tidak
82	Dani, D	P	855	930/890	1800	1720	Tidak	Tidak	Tidak
83	Eani, E	P	865	940/900	1820	1740	Tidak	Tidak	Tidak
84	Fani, F	P	875	950/910	1840	1760	Tidak	Tidak	Tidak
85	Gani, G	P	885	960/920	1860	1780	Tidak	Tidak	Tidak
86	Hani, H	P	895	970/930	1880	1800	Tidak	Tidak	Tidak
87	Iani, I	P	905	980/940	1900	1820	Tidak	Tidak	Tidak
88	Jani, J	P	915	990/950	1920	1840	Tidak	Tidak	Tidak
89	Kani, K	P	925	1000/960	1940	1860	Tidak	Tidak	Tidak
90	Lani, L	P	935	1010/970	1960	1880	Tidak	Tidak	Tidak
91	Mani, M	P	945	1020/980	1980	1900	Tidak	Tidak	Tidak
92	Nani, N	P	955	1030/990	2000	1920	Tidak	Tidak	Tidak
93	Oani, O	P	965	1040/1000	2020	1940	Tidak	Tidak	Tidak
94	Pani, P	P	975	1050/1010	2040	1960	Tidak	Tidak	Tidak
95	Qani, Q	P	985	1060/1020	2060	1980	Tidak	Tidak	Tidak
96	Rani, R	P	995	1070/1030	2080	2000	Tidak	Tidak	Tidak
97	Sani, S	P	1005	1080/1040	2100	2020	Tidak	Tidak	Tidak
98	Tani, T	P	1015	1090/1050	2120	2040	Tidak	Tidak	Tidak
99	Uani, U	P	1025	1100/1060	2140	2060	Tidak	Tidak	Tidak
100	Vani, V	P	1035	1110/1070	2160	2080	Tidak	Tidak	Tidak

Dataset yang digunakan dalam penelitian ini merupakan data pasien stroke yang tersimpan dalam format Microsoft Excel. Dataset tersebut berisi berbagai atribut klinis dan demografis yang berkaitan dengan faktor risiko stroke, seperti usia, jenis kelamin, status hipertensi, riwayat penyakit jantung, kadar

glukosa darah, dan indeks massa tubuh (BMI). Setiap baris data merepresentasikan satu individu pasien, sedangkan setiap kolom menggambarkan karakteristik tertentu dari pasien tersebut. Penggunaan data klinis dan demografis dalam analisis klastering terbukti efektif untuk mengidentifikasi kelompok pasien dengan karakteristik risiko yang serupa pada penelitian kesehatan, khususnya pada kasus stroke (Ren & Fu, 2021).

2. Preprocessing Data



Tahap preprocessing dilakukan untuk memastikan kualitas dataset sebelum proses klastering. Proses ini meliputi pemeriksaan terhadap nilai kosong (*missing values*), duplikasi data, serta inkonsistensi tipe data yang dapat memengaruhi hasil analisis. Pemeriksaan awal dilakukan menggunakan fitur eksplorasi data pada RapidMiner. Selanjutnya, atribut numerik dinormalisasi menggunakan operator *Normalize* agar seluruh variabel memiliki skala yang sebanding, sedangkan atribut kategorikal seperti jenis kelamin dan status penyakit dikonversi menjadi bentuk numerik menggunakan operator *Nominal to Numerical*. Tahap ini penting karena algoritma K-Means hanya dapat bekerja secara optimal pada data numerik berbasis jarak (Suhito et al., 2025).

3. Pembagian Data

Dataset yang telah diproses kemudian dibagi menjadi dua bagian, yaitu data pelatihan dan data pengujian. Pembagian dilakukan menggunakan operator *Split Data* dengan proporsi 90% data pelatihan dan 10% data pengujian. Pembagian ini bertujuan untuk memastikan algoritma memperoleh data yang cukup dalam membentuk cluster yang stabil sekaligus menjaga validitas proses evaluasi. Pembagian data yang tepat berperan penting dalam meningkatkan reliabilitas hasil analisis klastering serta meminimalkan bias penelitian (Jain, 2010).

4. Proses Klastering K-Means

Proses klastering dilakukan menggunakan operator *Clustering (K-Means)* pada RapidMiner. Algoritma K-Means bekerja dengan mengelompokkan data ke dalam sejumlah cluster berdasarkan jarak Euclidean antara data dan pusat cluster (*centroid*). Jumlah cluster ditentukan berdasarkan karakteristik dataset dan tujuan penelitian. Metode K-Means banyak digunakan dalam bidang kesehatan karena mampu mengelompokkan data medis

berskala besar dan menghasilkan pola yang mudah diinterpretasikan oleh peneliti maupun tenaga kesehatan (Shin et al., n.d.).

5. Evaluasi Hasil Klustering

Tahap evaluasi dilakukan untuk menilai kualitas hasil klustering yang diperoleh. Evaluasi dilakukan menggunakan operator *Performance (Clustering)* pada RapidMiner untuk mengukur tingkat pemisahan antar cluster dan tingkat keseragaman data dalam satu cluster. Penilaian kualitas kluster mengacu pada indeks evaluasi kluster, seperti Davies–Bouldin Index, yang digunakan untuk memastikan bahwa cluster yang terbentuk memiliki pemisahan yang jelas dan tidak saling tumpang tindih (Davies & Bouldin, 1979).

HASIL DAN PEMBAHASAN

Hasil Klustering Data Pasien Stroke



Berdasarkan tahapan metode penelitian yang telah dilakukan, proses klustering terhadap data pasien stroke berhasil menghasilkan beberapa kelompok pasien dengan karakteristik risiko yang berbeda. Penerapan algoritma K-Means pada dataset yang telah melalui tahap preprocessing dan normalisasi memungkinkan pengelompokan data berdasarkan tingkat kemiripan atribut klinis dan demografis pasien, seperti usia, status hipertensi, kadar glukosa darah, dan indeks massa tubuh (BMI).

Hasil klustering menunjukkan bahwa setiap cluster memiliki pola karakteristik yang relatif homogen di dalam cluster dan heterogen antar cluster. Cluster tertentu didominasi oleh pasien dengan usia lanjut, kadar glukosa darah tinggi, serta riwayat hipertensi dan penyakit jantung, yang mengindikasikan kelompok dengan risiko stroke yang lebih tinggi. Sementara itu, cluster lainnya menunjukkan karakteristik pasien dengan usia lebih muda, nilai BMI normal, serta kadar glukosa darah yang relatif stabil, sehingga dapat diinterpretasikan sebagai kelompok dengan risiko stroke yang lebih rendah.

Distribusi data pada setiap cluster menunjukkan bahwa algoritma K-Means mampu membagi dataset secara proporsional tanpa menghasilkan cluster yang terlalu kecil atau terlalu dominan. Hal ini

menandakan bahwa struktur data memiliki pola alami yang dapat dieksplorasi secara efektif melalui pendekatan klastering.

Evaluasi Kualitas Klastering

PerformanceVector

```
PerformanceVector:  
Avg. within centroid distance: 4.919  
Avg. within centroid distance_cluster_0: 3.781  
Avg. within centroid distance_cluster_1: 4.733  
Avg. within centroid distance_cluster_2: 10.832  
Avg. within centroid distance_cluster_3: 5.409  
Avg. within centroid distance_cluster_4: 7.498  
Davies Bouldin: 1.431
```

Evaluasi kualitas hasil klastering dilakukan menggunakan operator *Performance (Clustering)* pada RapidMiner. Evaluasi ini bertujuan untuk menilai sejauh mana cluster yang terbentuk memiliki pemisahan yang jelas dan tingkat keseragaman yang baik dalam setiap cluster. Salah satu indikator utama yang digunakan adalah *Davies–Bouldin Index (DBI)*, di mana nilai yang lebih rendah menunjukkan kualitas klastering yang lebih baik.

Hasil evaluasi menunjukkan bahwa nilai indeks klaster yang diperoleh berada pada rentang yang dapat diterima, yang menandakan bahwa jarak antar cluster cukup besar dibandingkan dengan jarak antar data di dalam satu cluster. Dengan demikian, cluster yang terbentuk tidak saling tumpang tindih secara signifikan dan memiliki karakteristik yang cukup jelas untuk dibedakan.

Selain itu, hasil evaluasi juga menunjukkan bahwa normalisasi data berperan penting dalam meningkatkan kualitas klastering. Tanpa normalisasi, atribut dengan skala besar seperti kadar glukosa darah berpotensi mendominasi proses pembentukan cluster, sedangkan setelah normalisasi, setiap atribut memberikan kontribusi yang seimbang dalam perhitungan jarak Euclidean.

Pembahasan

Hasil penelitian ini menunjukkan bahwa metode klastering K-Means mampu mengidentifikasi pola risiko stroke secara efektif berdasarkan data klinis dan demografis pasien. Pengelompokan yang dihasilkan memberikan gambaran bahwa faktor-faktor seperti usia, hipertensi, riwayat penyakit jantung, kadar glukosa darah, dan BMI memiliki keterkaitan yang kuat dalam membentuk kelompok risiko stroke tertentu.

Pendekatan klastering memungkinkan analisis dilakukan secara tidak terawasi (unsupervised), sehingga sangat bermanfaat dalam kondisi di mana label risiko tidak tersedia atau belum terdefinisi secara jelas. Dengan mengandalkan kemiripan data, K-Means mampu mengungkap struktur tersembunyi dalam dataset dan memberikan wawasan awal mengenai segmentasi pasien berdasarkan tingkat risiko.

Dari sisi praktis, hasil klastering ini dapat dimanfaatkan sebagai dasar dalam pengambilan keputusan di bidang kesehatan, khususnya untuk mendukung strategi pencegahan stroke. Tenaga medis dapat menggunakan informasi cluster untuk memfokuskan perhatian pada kelompok pasien dengan risiko tinggi melalui program pemantauan intensif, edukasi kesehatan, dan intervensi dini. Sementara itu, kelompok dengan risiko lebih rendah dapat diarahkan pada strategi pencegahan jangka panjang.

Secara keseluruhan, hasil dan pembahasan ini menegaskan bahwa metode K-Means berbasis RapidMiner merupakan pendekatan yang efektif dan efisien dalam menganalisis pola risiko stroke. Metode ini tidak hanya mampu mengelompokkan pasien berdasarkan karakteristik yang relevan, tetapi juga memberikan interpretasi yang mudah dipahami dan dapat dijadikan landasan awal dalam pengembangan sistem pendukung keputusan berbasis data di bidang kesehatan.

KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan, dapat disimpulkan bahwa metode klastering K-Means mampu mengidentifikasi dan menganalisis pola risiko stroke pada pasien secara efektif menggunakan data klinis dan demografis. Melalui tahapan penelitian yang sistematis, mulai dari pengumpulan dataset, preprocessing, normalisasi data, hingga proses klastering dan evaluasi, penelitian ini berhasil membentuk kelompok pasien yang memiliki karakteristik risiko yang relatif homogen dalam satu cluster dan heterogen antar cluster.

Hasil klastering menunjukkan adanya perbedaan pola yang jelas antar kelompok pasien, di mana cluster tertentu didominasi oleh pasien dengan faktor risiko tinggi seperti usia lanjut, kadar glukosa darah yang tinggi, serta riwayat hipertensi dan penyakit jantung, sedangkan cluster lainnya menunjukkan karakteristik risiko yang lebih rendah. Evaluasi kualitas klastering menggunakan metrik evaluasi clustering menunjukkan bahwa cluster yang terbentuk memiliki tingkat pemisahan yang baik dan tidak saling tumpang tindih secara signifikan.

Dengan demikian, penelitian ini membuktikan bahwa pendekatan klastering K-Means berbasis RapidMiner dapat digunakan sebagai metode analisis yang efektif untuk menggali pola risiko stroke secara tidak terawasi. Temuan ini diharapkan dapat menjadi dasar awal dalam mendukung pengambilan keputusan di bidang kesehatan, khususnya dalam perencanaan strategi pencegahan dan pemantauan pasien stroke berbasis data.

DAFTAR PUSTAKA

- Care, P., Delord, M., Sun, X., Learoyd, A., Curcin, V., Wolfe, C., Ashworth, M., & Douiri, A. (2024). Patient - oriented unsupervised learning to uncover the patterns of multimorbidity associated with stroke using primary care electronic health records. *BMC Primary Care*. <https://doi.org/10.1186/s12875-024-02636-6>
- Ceskoutsé, R. F. T., Bomgni, A. B., Zanfack, D. R. G., Agany, D. D. M., Thomas, B. B., & Zohim, E. G. (2025). *Sub-clustering based recommendation system for stroke patient: Identification of a specific drug class for a given patient*. 1–35. <https://doi.org/10.1016/j.compbimed.2024.108117>. Sub-clustering
- Davies, D. L., & Bouldin, D. W. (1979). A Cluster Separation Measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-1(2), 224–227.

- <https://doi.org/10.1109/TPAMI.1979.4766909>
- Guo, X., Liu, P., Guo, J., Zhang, N., Huang, H., Liu, J., Tan, Z., & Dan, G. (2025). *An unsupervised cluster analysis of multimorbidity patterns in older adults in Shenzhen , China*. June, 1–14. <https://doi.org/10.3389/fpubh.2025.1557721>
- Id, S. S. Z., Rutter, M. K., Sun, L. Y., Ullah, W., Rashid, M., Id, D. M. A., Steinke, D. T., Weng, S., Id, E. K., & Mamas, M. A. (2023). *PLOS ONE Comorbidity clusters and in-hospital outcomes in patients admitted with acute myocardial infarction in the USA : A national population-based study*. 1–18. <https://doi.org/10.1371/journal.pone.0293314>
- Jain, A. K. (2010). Data clustering: 50 years beyond K-means. *Pattern Recognition Letters*, 31(8), 651–666. <https://doi.org/https://doi.org/10.1016/j.patrec.2009.09.011>
- Jiang, Y., Dang, Y., Wu, Q., & Yuan, B. (n.d.). *Using a k-means clustering to identify novel phenotypes of acute ischemic stroke and development of its Clinlabomics models*.
- K-means, A. (2024). *Clustering Pasien Rawat Inap Di RS USU Menggunakan*. 03(2), 54–63.
- Kim, J. T., Kim, N. R., Choi, S. H., Oh, S., Park, M. S., Lee, S. H., Kim, B. C., Choi, J., & Kim, M. S. (2022). Neural network - based clustering model of ischemic stroke patients with a maximally distinct distribution of 1 - year vascular outcomes. *Scientific Reports*, 0123456789, 1–10. <https://doi.org/10.1038/s41598-022-13636-w>
- Mao, H. (2025). *Cluster Analysis of Patients With Acute Ischemic Stroke : Identifying Characteristics of Long Hospital Stays in a Comprehensive Hospital*. 1–10. <https://doi.org/10.1002/brb3.70940>
- Ren, Z., & Fu, X. (2021). Stroke Risk Factors in United States: An Analysis of the 2013-2018 National Health and Nutrition Examination Survey. *International Journal of General Medicine*, 14, 6135–6147. <https://doi.org/10.2147/IJGM.S327075>
- Rumah, P., Royal, S., Hidayah, A., Dulisep, D., & Angga, B. (2024). *Implementasi Algoritma K - Means Menggunakan RapidMiner untuk Klasterisasi Data Obat*. 7(2), 200–211.
- Shin, S., Chang, W. H., Kim, D. Y., Lee, J., Sohn, M. K., Song, M., Shin, Y., Lee, Y., Joo, M. C., Lee, S. Y., Han, J., Ahn, J., Oh, G., Kim, Y., Kim, K., & Kim, Y. (n.d.). *Clustering and prediction of long-term functional recovery patterns in first-time stroke patients*. <https://doi.org/10.30597/mkmi.v21i3.45641>
- Suhito, H. P., Azam, M., Nur, D., Ningrum, A., City, S., & Office, H. (2025). *Media Kesehatan Masyarakat Indonesia Indonesia : Based on Indonesian Health Survey Data*. 21(3), 225–235. <https://doi.org/10.30597/mkmi.v21i3.45641>
- Wala, J., & Umar, R. (2024). *Implementasi K-Means Clustering pada Pengelompokan Pasien Penyakit Jantung*. 9(3), 205–216. <https://doi.org/10.14421/jiska.2024.9.3.205-216>