



PENGELOMPOKAN PENDERITA GANGGUAN TIDUR BERDASARKAN GAYA HIDUP MENGUNAKAN ALGORITMA K-MEANS CLUSTERING

Bagas Wira Yuda^{a*}, Zaehol Fatah^b

^a Sistem Informasi, wirabagas60@gmail.com, Universitas Ibrahimy, Situbondo Jawa Timur

^b Sistem Informasi, zaeholfatah@gmail.com, Universitas Ibrahimy, Situbondo Jawa Timur

* korespondensi

ABSTRACT

Sleep disorders, including insomnia, can be influenced by various lifestyle factors, such as sleep duration, sleep quality, physical activity, and individual health conditions. This study aims to categorize the risk level of insomnia based on lifestyle using the K-Means clustering algorithm. The data used include sleep duration, sleep quality, heart rate, and daily step count. Through the implementation of the K-Means algorithm, the data is analyzed to group individuals into several categories based on existing lifestyle patterns. The results of the study show a correlation between a healthy lifestyle and better sleep quality. In addition, the resulting clusters provide insight into lifestyle characteristics that affect the risk of insomnia, so that they can be the basis for recommendations for more targeted health interventions. This study is expected to contribute to the development of data-based sleep disorder management strategies by utilizing machine learning methods, especially the K-Means algorithm, to support efforts to improve the quality of life of the community.

Keywords: Data Mining, Information Systems, Sleep Disorders, Insomnia.

Abstrak

Gangguan tidur, termasuk insomnia, dapat dipengaruhi oleh berbagai faktor gaya hidup, seperti durasi tidur, kualitas tidur, aktivitas fisik, dan kondisi kesehatan individu. Penelitian ini bertujuan untuk mengkategorikan tingkat risiko insomnia berdasarkan gaya hidup menggunakan algoritma clustering K-Means. Data yang digunakan meliputi durasi tidur, kualitas tidur, denyut jantung, dan jumlah langkah harian. Melalui penerapan algoritma K-Means, data dianalisis untuk mengelompokkan individu ke dalam beberapa kategori berdasarkan pola gaya hidup yang ada. Hasil penelitian menunjukkan adanya korelasi antara gaya hidup sehat dengan kualitas tidur yang lebih baik. Selain itu, cluster yang dihasilkan memberikan wawasan tentang karakteristik gaya hidup yang memengaruhi risiko insomnia, sehingga dapat menjadi dasar rekomendasi intervensi kesehatan yang lebih terarah. Penelitian ini diharapkan dapat memberikan kontribusi bagi pengembangan strategi manajemen gangguan tidur berbasis data dengan memanfaatkan metode machine learning, khususnya algoritma K-Means, untuk mendukung upaya peningkatan kualitas hidup masyarakat.

Kata Kunci: Data Mining, Sistem Informasi, Gangguan Tidur, Insomnia.

1. PENDAHULUAN

Kemajuan teknologi kedokteran telah membuka peluang besar untuk mengatasi berbagai tantangan dalam mendiagnosis dan mengkategorikan kondisi kesehatan, termasuk gangguan tidur. Gangguan tidur, seperti insomnia, sleep apnea, dan lainnya merupakan kondisi medis yang memengaruhi kemampuan seseorang untuk mendapatkan kualitas atau kuantitas tidur yang cukup. Gangguan tersebut dapat ditandai dengan berbagai gejala, seperti sulit tidur, sering terbangun di malam hari, merasa lelah meskipun sudah cukup tidur, atau pola tidur tidak teratur sehingga memengaruhi fungsi tubuh sehari-hari. Menurut survei yang

dilakukan oleh National Sleep Foundation (NSF) di Amerika, orang-orang yang mengaku mengalami kesulitan tidur dilaporkan mengalami lebih banyak kesulitan berkonsentrasi, kesulitan menyelesaikan sejumlah tugas, serta sedikit emosional atau mudah tersinggung.

Salah satu pendekatan yang dapat digunakan adalah penerapan algoritma K-Means Clustering, yaitu metode data mining yang bertujuan untuk mengelompokkan data ke dalam beberapa kelompok (cluster) berdasarkan karakteristik tertentu. Algoritma ini sangat cocok untuk menganalisis data yang besar dengan banyak variabel, seperti data gaya hidup dan pola tidur. Dengan algoritma ini, penderita gangguan tidur dapat dikelompokkan berdasarkan pola yang serupa, seperti tingkat aktivitas fisik, durasi tidur, atau konsumsi makanan tertentu. Hasil pengelompokan ini dapat membantu dalam memahami penyebab gangguan tidur dan merancang intervensi yang lebih tepat sasaran.

Dalam penerapannya, algoritma K-Means bekerja dengan mengelompokkan individu berdasarkan data numerik yang relevan. Misalnya, durasi tidur dan pola aktivitas harian dapat diolah menjadi dataset yang diukur dan dikelompokkan. Proses ini menghasilkan cluster yang mewakili kelompok dengan karakteristik yang serupa, misalnya cluster individu dengan durasi tidur yang sangat pendek, pola gaya hidup pasif, atau konsumsi kafein yang tinggi. Informasi ini sangat berguna bagi peneliti dan praktisi kesehatan dalam menentukan pendekatan yang dipersonalisasi untuk mengatasi gangguan tidur pada setiap kelompok.

Penerapan K-Means Clustering dalam konteks gangguan tidur berbasis gaya hidup memberikan banyak manfaat, antara lain efisiensi dalam analisis data, penyajian pola yang mudah dipahami, dan menjadi dasar pengembangan program pencegahan atau pengobatan yang lebih efektif. Selain itu, hasil clustering dapat menjadi dasar pengembangan kebijakan kesehatan masyarakat, seperti kampanye gaya hidup sehat yang disesuaikan dengan kebutuhan masing-masing kelompok. Dengan demikian, algoritma ini tidak hanya meningkatkan pemahaman ilmiah tentang gangguan tidur, tetapi juga memberikan dampak nyata dalam meningkatkan kualitas hidup individu yang mengalaminya.

2. TINJAUAN PUSTAKA

Gangguan tidur merupakan masalah kesehatan yang signifikan dan terus meningkat sehingga berdampak negatif pada kualitas hidup dan kesejahteraan individu. Gangguan tidur berkaitan erat dengan gaya hidup seseorang seperti durasi tidur sehari-hari, langkah harian, dan jumlah detak jantung yang mempengaruhi kualitas tidur. Gangguan tidur dapat mengurangi kenyamanan dan waktu tidur sehingga akan menyebabkan kualitas tidur yang buruk. Gangguan tidur yang seringkali terjadi adalah insomnia, sleep apnea, restless legs syndrome, hipersomnia, circadian rhythm disorders, dan parasomnia. Data pada penelitian ini bertujuan untuk dianalisis dan dikelompokkan berdasarkan gangguan tidur yang disebabkan oleh gaya hidup melalui salah satu metode data mining yaitu K-Means Clustering melalui identifikasi beberapa cluster yang memiliki gangguan tidur dan tidak memiliki gangguan tidur. Dengan menggunakan metode KMeans diperoleh bahwa berdasarkan data Sleep Duration dan Quality of Sleep terdapat 174 data pada cluster 1 yang menunjukkan adanya gangguan tidur insomnia, terdapat 164 data pada cluster 2 yang menunjukkan tidak adanya gangguan tidur, dan terdapat 36 data pada cluster 3 yang menunjukkan adanya gangguan tidur berupa sleep apnea dengan struktur silhouette score standar (medium structure). Sedangkan berdasarkan data Heart Rate dan Daily Steps terdapat 174 data pada cluster 1 yang menunjukkan adanya gangguan tidur insomnia, terdapat 164 data pada cluster 2 yang menunjukkan adanya gangguan tidur berupa sleep apnea, dan terdapat 36 data pada cluster 3 yang menunjukkan tidak adanya gangguan tidur dengan struktur silhouette score kuat (strong structure). (Implementasi Algoritma K-Means Clustering dalam Penentuan Gangguan Tidur Seseorang berdasarkan Gaya Hidup).

Perencanaan kebutuhan obat-obatan yang tepat akan membuat pengadaan obat-obatan menjadi efektif dan efisien sehingga tersedia obat-obatan dengan jenis dan jumlah yang cukup sesuai dengan kebutuhan serta dapat diperoleh pada saat yang diperlukan. Clustering dalam data mining dapat digunakan untuk menganalisa pemakaian obat-obatan, perencanaan dan pengendalian obat-obatan di rumah sakit. Metode yang akan di pakai untuk clustering data obat-obatan adalah algoritma K-Means yang merupakan metode data clustering non hirarki yang mempartisi data ke dalam cluster sehingga data yang memiliki karakteristik yang sama dikelompokkan ke dalam satu cluster yang sama dan data yang mempunyai karakteristik yang berbeda dikelompokkan ke dalam kelompok lain. Tujuan dari penelitian ini adalah untuk mengelompokkan data obat-obatan di Rumah Sakit Umum Daerah Pekanbaru yang dapat digunakan sebagai referensi dalam pengambilan keputusan dalam perencanaan dan pengendalian pasokan medis di rumah sakit tersebut. (Penerapan Algoritma K-Means Untuk Clustering Data Obat-Obatan Pada RSUD Pekanbaru)

3. METODOLOGI PENELITIAN

Penelitian ini berfokus pada penerapan metode clustering menggunakan algoritma K-means untuk mengelompokkan penderita gangguan tidur berdasarkan gaya hidup.



Gambar 1 Metode Penelitian

Tahap pertama dari penelitian ini adalah pengumpulan data yang dilakukan untuk menyediakan informasi mentah yang akan diolah lebih lanjut. Setelah data terkumpul, langkah selanjutnya adalah Pengolahan Data yang merupakan tahap krusial untuk menyiapkan data. Selanjutnya, data yang diolah masuk ke tahap inti yaitu Clustering. Pada tahap ini, algoritma K-Means digunakan untuk mengelompokkan data berdasarkan karakteristik yang serupa. Tahap terakhir adalah Result, di mana hasil clustering divisualisasikan dan diinterpretasikan.

3.1 Pengumpulan Data

Pengumpulan data dalam pengelompokan gangguan tidur merupakan tahap awal dalam proses analisis data yang bertujuan untuk memperoleh informasi yang relevan terkait pola tidur dan gaya hidup individu. Data yang dikumpulkan biasanya mencakup variabel seperti durasi tidur, tingkat aktivitas fisik (seperti langkah harian), detak jantung, dan kebiasaan lain yang memengaruhi kualitas tidur.

3.2 Pembersihan Data

Pembersihan data dalam clustering gangguan tidur adalah proses pembersihan dan penyiapan data agar sesuai untuk dianalisis. Tahap ini meliputi pengecekan, penghapusan, atau perbaikan data yang tidak valid, hilang, atau tidak konsisten. Proses ini penting untuk memastikan bahwa data yang digunakan dalam algoritma clustering, seperti K-Means, tidak mengandung *noise* atau *error* yang dapat memengaruhi akurasi hasil *clustering*.

3.3 Clustering

Clustering merupakan salah satu teknik analisis data yang bertujuan untuk mengelompokkan individu dengan pola atau karakteristik yang serupa ke dalam kelompok tertentu. Pendekatan ini tidak memerlukan data berlabel dan sering digunakan untuk mengidentifikasi pola tersembunyi dalam big data, seperti pola tidur, durasi tidur, atau faktor yang berhubungan dengan gangguan tidur.

3.4 K-Means

K-Means merupakan salah satu algoritma clustering yang digunakan untuk mengelompokkan data ke dalam beberapa kelompok atau cluster berdasarkan kemiripan antar data. Algoritma ini bekerja dengan membagi data ke dalam k kelompok, dimana setiap kelompok memiliki pusat cluster (centroid). Proses ini dilakukan secara iteratif untuk memperkecil jarak antara data dengan centroid kelompok.

3.5 Davies Bouldin

Indeks Davies-Bouldin (DBI) merupakan metrik evaluasi yang digunakan untuk mengukur kualitas hasil dari suatu algoritma clustering. DBI dihitung berdasarkan rata-rata kemiripan setiap cluster dengan cluster lainnya. Semakin kecil nilai DBI, maka hasil clustering akan semakin baik, karena menunjukkan cluster yang terbentuk semakin terpisah secara signifikan dan memiliki unsur yang lebih homogen. DBI diperkenalkan oleh David L. Davies dan Donald W. Bouldin pada tahun 1979. Metode ini merupakan

metode evaluasi internal, artinya pengujian dilakukan dengan menggunakan karakteristik data itu sendiri tanpa memerlukan data eksternal sebagai acuan.

3.6 RapidMiner

RapidMiner adalah platform perangkat lunak yang digunakan untuk ilmu data dan penambangan data. RapidMiner menyediakan lingkungan terpadu untuk persiapan data, pemodelan, evaluasi, dan penyebaran hasil analisis data. Perangkat lunak ini memanfaatkan berbagai teknik pembelajaran mesin untuk menganalisis data dan membantu pengguna dalam pengambilan keputusan berdasarkan data. RapidMiner sering digunakan dalam pendidikan, penelitian, dan bisnis untuk menganalisis big data dan membantu berbagai kasus seperti prediksi, pengelompokan, dan pengelompokan data.

4. HASIL DAN PEMBAHASAN

4.1 Pengumpulan Data

Tahap pertama, Pengumpulan Data, merupakan dasar dari keseluruhan proses. Pengumpulan data dilakukan untuk menyediakan informasi mentah yang akan diolah lebih lanjut. Dataset yang digunakan dalam penelitian ini berasal dari sumber terbuka, yaitu Kaggle, dengan nama Sleep Health and Lifestyle Dataset. Data yang digunakan dalam dataset tersebut meliputi Person ID, Usia, Durasi Tidur, Kualitas Tidur, Denyut Jantung, dan Langkah Harian. Tabel di bawah ini menunjukkan dataset awal penelitian. Data yang diolah sebanyak 149 data, sementara hanya sebagian saja yang disertakan dalam tabel ini:

Tabel 1 Dataset Awal

Person ID	Age	Sleep Duration	Quality of Sleep	Heart Rate	Daily Steps
1	27	6.1	6	77	4200
2	28	6.2	6	75	10000
3	28	6.2	6	75	10000
4	28	5.9	4	85	3000
5	28	5.9	4	85	3000
6	28	5.9	4	85	3000
7	29	6.3	6	82	3500
8	29	7.8	7	70	8000
9	29	7.8	7	70	8000
10	29	7.8	7	70	8000
11	29	6.1	6	70	8000
12	29	7.8	7	70	8000
13	29	6.1	6	70	8000
14	29	6	6	70	8000

Sumber: Kaggle

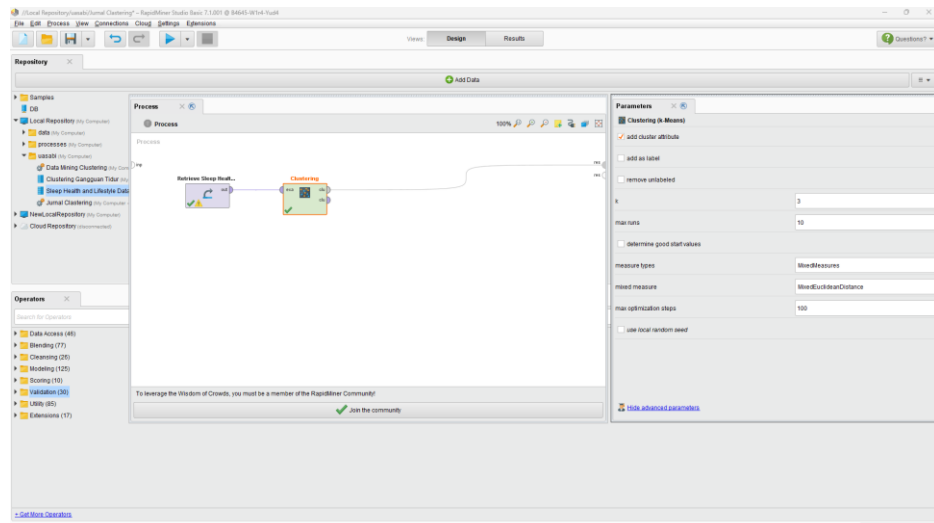
4.2 Pembersihan Data

Pembersihan data merupakan proses perbaikan atau penghapusan data yang salah, rusak, salah format, duplikat, atau tidak lengkap pada suatu kumpulan data. Proses ini dilakukan untuk mendapatkan data yang bersih dan berkualitas sehingga proses dapat berlanjut ke tahap selanjutnya yaitu Clustering. Aplikasi Microsoft Excel sering digunakan dalam pembersihan data karena tools yang disediakan lebih mudah digunakan.

4.3 Clustering

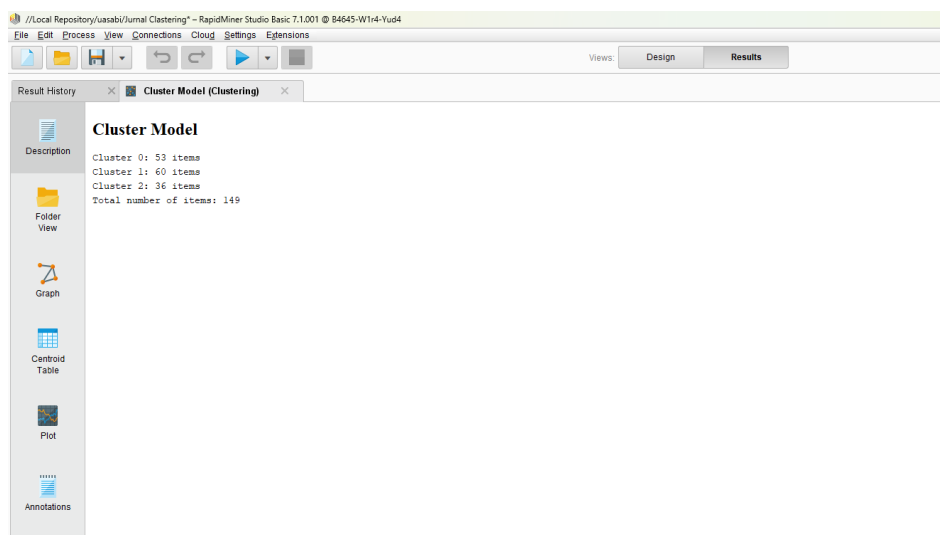
Pada tahap ini dilakukan proses penerapan data mining. Dengan metode algoritma K-Means, penulis mengelompokkan data penderita gangguan tidur berdasarkan gaya hidup. Proses algoritma K-Means adalah sebagai berikut:

- Langkah pertama yang dilakukan adalah memasukkan dataset penderita gangguan tidur ke dalam Aplikasi RapidMiner. Dataset harus berupa soft file Microsoft Excel dengan format .csv seperti pada tabel Dataset Awal di atas.
- Setelah data dibersihkan dan diinput, penulis membuat model clustering dengan metode K-Means yang kemudian disusun sesuai cluster yang diinginkan. Di sini penulis mengelompokkan dataset menjadi 3 cluster dengan 10 iterasi. Operator K-Means dapat dilihat pada gambar di bawah ini:



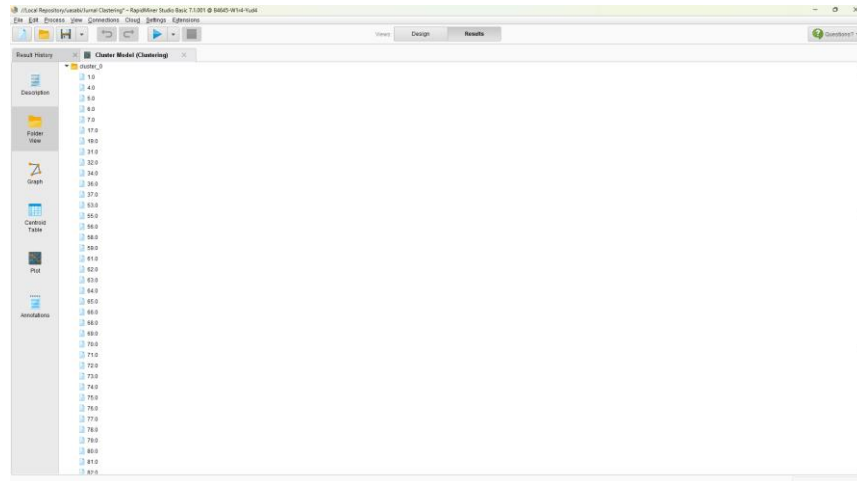
Gambar 2 Operator K-Means

- c. Setelah dilakukan pengujian menggunakan operator K-Means pada aplikasi RapidMiner, dari 149 data didapatkan hasil pengelompokan penderita gangguan tidur yang terdiri dari 3 cluster yaitu cluster 0, cluster 1, dan cluster 2. Berikut ini merupakan hasil cluster dari dataset penderita gangguan tidur:
- d.

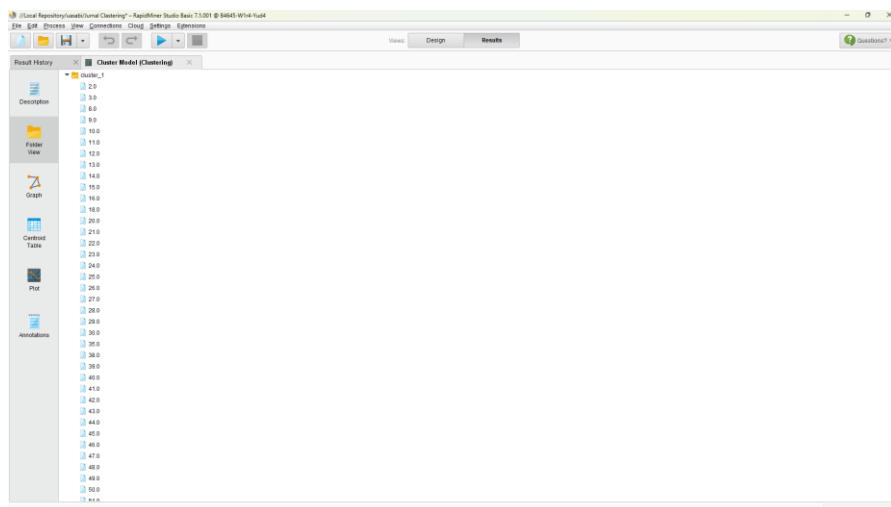


Gambar 3 Hasil Clustering

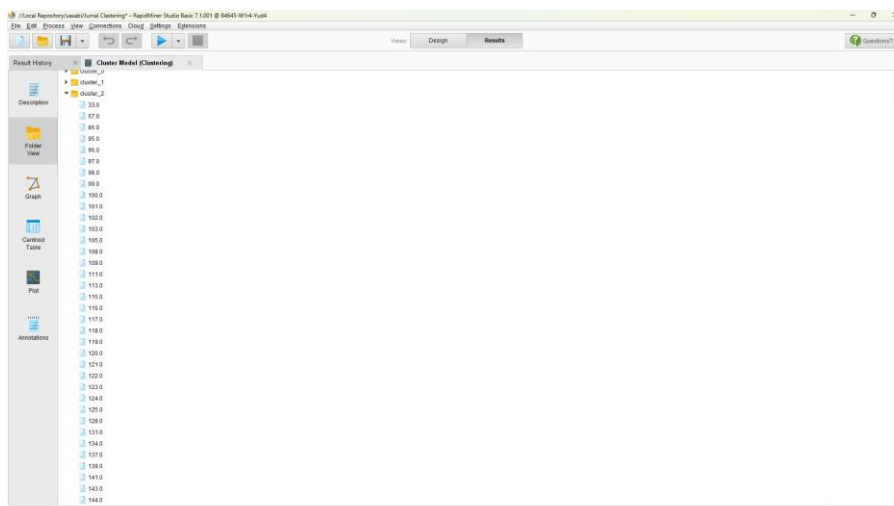
- e. Berikut ini adalah tampilan anggota masing-masing cluster yang telah diuji menggunakan algoritma K-Means pada aplikasi RapidMiner:



Gambar 4 Cluster_0

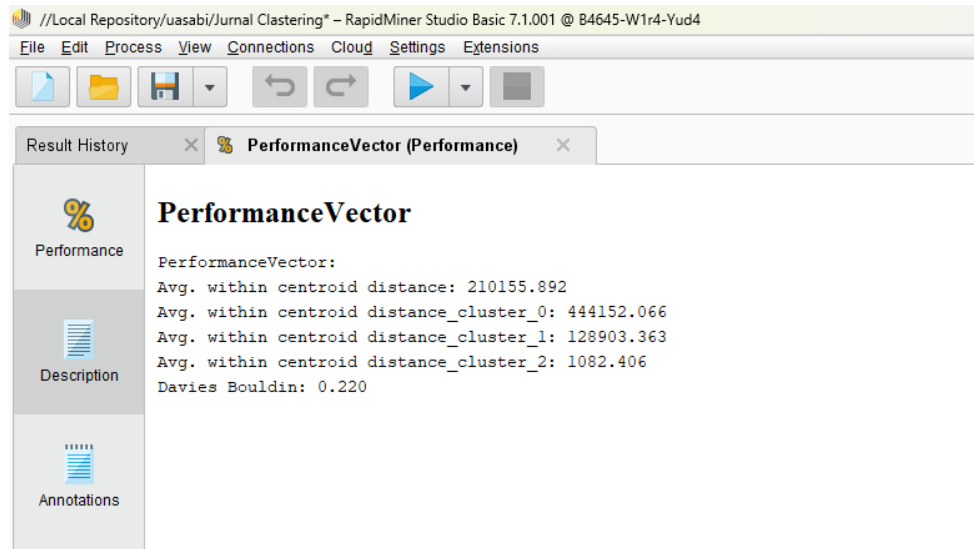


Gambar 5 Cluster_1



Gambar 6 Cluster_2

- f. Terakhir, hitung nilai Davies Bouldin pada cluster di aplikasi RapidMiner menggunakan operator Cluster Distance Performance untuk mengetahui seberapa efektif cluster tersebut dan mengevaluasi cluster secara umum. Berikut ini adalah hasil perhitungan nilai Davies Bouldin:



Gambar 7 Davies Bouldin

4.4 Implementasi

Dari 149 data penderita gangguan tidur yang telah diuji dengan aplikasi RapidMiner, diperoleh hasil sebagai berikut: Cluster 0 memiliki 53 anggota, cluster 1 memiliki 60 anggota, dan cluster 2 memiliki 36 anggota. Performance Vector-nya adalah sebagai berikut:

- a. *Avg. within centroid distance: 210155.892*
- b. *Avg. within centroid distance_cluster_0: 444152.066*
- c. *Avg. within centroid distance_cluster_1: 128903.363*
- d. *Avg. within centroid distance_cluster_2: 1082.406*
- e. *Davies Bouldin: 0.220*

Berdasarkan hasil pengelompokan di atas, nilai Average. pada jarak pusat massa menyentuh angka 210155.892. Angka ini menunjukkan jarak rata-rata setiap data di semua cluster terhadap centroid-nya. Semakin kecil nilai ini, semakin baik karena menunjukkan bahwa data dalam cluster tersebut semakin dekat dengan pusat cluster, yang mencerminkan kohesi yang tinggi. Nilai 210155.892 menunjukkan bahwa jarak keseluruhan masih cukup besar, yang mungkin menunjukkan beberapa cluster tidak terlalu kompak. Berikut analisisnya:

4.4.1. Cluster 0 (444152.066):

Cluster ini memiliki jarak rata-rata tertinggi dari centroid, yang menunjukkan bahwa data dalam cluster ini lebih menyebar daripada cluster lainnya. Ini menunjukkan bahwa hasil cluster mencakup pasien dengan pola atau gejala tidur yang sangat bervariasi.

4.4.2. Cluster 1 (128903.363):

Cluster ini memiliki sebaran yang lebih kecil daripada cluster 0, tetapi masih signifikan. Ini menunjukkan adanya kelompok dengan karakteristik gangguan tidur yang lebih seragam.

4.4.3. Cluster 2 (1082.406):

Cluster ini memiliki jarak rata-rata terkecil, yang menunjukkan bahwa data dalam cluster ini sangat dekat dengan centroid. Kemungkinan besar, ini adalah kelompok dengan pola atau gejala yang sangat seragam. Hasil ini dapat digunakan untuk mengembangkan intervensi atau kelompok perawatan yang lebih spesifik untuk orang dengan gangguan tidur, misalnya dengan fokus utama pada Cluster 2 untuk perawatan yang lebih seragam.

5. KESIMPULAN DAN SARAN

Hasil clustering data penderita gangguan tidur menunjukkan bahwa data dapat dikelompokkan menjadi tiga cluster dengan karakteristik yang berbeda berdasarkan jarak rata-rata masing-masing data terhadap pusat clusternya. Cluster 0 memiliki jarak rata-rata tertinggi (444152.066), yang menunjukkan bahwa data dalam cluster ini sangat tersebar atau memiliki tingkat variasi yang tinggi. Hal ini dapat menunjukkan bahwa cluster ini mencakup penderita dengan pola gangguan tidur yang beragam, kemungkinan terdiri dari

berbagai jenis gangguan atau kondisi yang memengaruhi tidur secara kompleks. Sebaliknya, Cluster 2 memiliki jarak rata-rata terendah (1082.406), yang menunjukkan bahwa data dalam cluster ini sangat padat dan seragam. Kemungkinan besar, cluster ini berisi penderita dengan pola atau gejala gangguan tidur yang sangat mirip, sehingga dapat menjadi target utama untuk pendekatan penanganan yang seragam.

Selain itu, nilai Davies Bouldin Index (DBI) sebesar 0,220 menunjukkan bahwa hasil clustering memiliki kualitas yang sangat baik. Cluster yang dihasilkan memiliki tingkat pemisahan yang tinggi dan distribusi yang kompak. Hal ini menunjukkan bahwa metode clustering telah berhasil memisahkan data penderita gangguan tidur ke dalam kelompok yang bermakna. Dengan hasil ini, para pemangku kepentingan, seperti dokter atau peneliti, dapat menggunakan pembagian cluster ini untuk merancang intervensi yang lebih spesifik. Misalnya, Cluster 2 dapat menjadi prioritas untuk solusi yang lebih homogen, sedangkan Cluster 0 memerlukan analisis lebih lanjut karena keragaman data di dalamnya.

DAFTAR PUSTAKA

- [1] S. F. Handoko, "IMPLEMENTASI DATA MINING UNTUK MENENTUKAN TINGKAT PENJUALAN PAKET DATA TELKOMSEL MENGGUNAKAN METODE K-MEANS CLUSTERING," *Jurnal Ilmiah Teknologi dan Rekayasa*, vol. 25, no. 1, 2020.
- [2] E. F. H. A. Maulida, "Implementasi Algoritma K-Means Clustering dalam Penentuan Gangguan Tidur Seseorang berdasarkan Gaya Hidup," 2024.
- [3] A. M. H. D. Y. Gustientiedina, "Penerapan Algoritma K-Means Untuk Clustering Data Obat-Obatan," *Jurnal Nasional Teknologi dan Sistem Informasi*, vol. 5, no. 1, 2019.
- [4] S. M. N. T. Butsiyanto, "Penerapan Data Mining Untuk Prediksi Penjualan Mobil Menggunakan Metode K-Means Clustering," *Jurnal Nasional Komputasi dan Teknologi Informasi*, vol. 3, 2020.
- [5] T. B. D. P. A. S. B. N. Amalina, "Metode K-Means Clustering Dalam Pengelompokan Penjualan Produk Frozen Food," *Jurnal Ilmiah Wahana Pendidikan*, vol. 8, no. 15, 2022.
- [6] K. Handoko, "PENERAPAN DATA MINING DALAM MENINGKATKAN MUTU PEMBELAJARAN PADA INSTANSI PERGURUAN TINGGI MENGGUNAKAN METODE K-MEANS CLUSTERING (STUDI KASUS DI PROGRAM STUDI TKJ AKADEMI KOMUNITAS SOLOK SELATAN)," *TEKNOSI*, vol. 02, 2016.
- [7] S. F. P. L. A. F. Woro Isti Rahayu, "IMPLEMENTASI DATA MINING DENGAN METODE K-MEANS CLUSTERING UNTUK MENENTUKAN IKLAN AUDIO BERDASARKAN USER BEHAVIORS PADA APLIKASI AUDIO SOCIAL MEDIA SVARA DI PT. ZAMRUD KHATULISTIWA TECHNOLOGY," *Jurnal Teknik Informatika*, vol. 10, 2018.
- [8] T. Hidayat, "Klasifikasi Data Jamaah Umroh Menggunakan Metode K-Means Clustering," *Jurnal Sistim Informasi dan Teknologi*, 2022.
- [9] R. S. Wahono, "Data Mining," 2020.
- [10] T. Suprawoto, "KLASIFIKASI DATA MAHASISWA MENGGUNAKAN METODE K-MEANS UNTUK MENUNJANG PEMILIHAN STRATEGI PEMASARAN," vol. 1, 2016.
- [11] I. C. Nisa, "PENERAPAN ALGORITMA K-MEANS DALAM PENGKATEGORIAN INSOMNIA," *METHOMIKA Jurnal Manajemen Informatika dan Komputerisasi Akuntansi*, 2024.
- [12] P. H. T. S. Syahriani, "Analisis Clustering Menggunakan Algoritma K-Means Dalam Pengelompokan Penjualan Produk Bahan Bangunan," *Journal of Information System Research*, vol. 6, no. 1, 2024.