



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: 1168–1175

Analisis Clustering Data Produk Menggunakan Algoritma *Self Organizing Map*

Dwi Ramadhani Prasetyo, Hasbi Firmansyah, Wahyu Asriyani

Universitas Pancasakti Tegal

Article in Journal of MISTER

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

DOI : <https://doi.org/10.32672/mister.v3i1.4043>

How to Cite this Article

APA : Prasetyo, D. R., Hasbi Firmansyah, & Wahyu Asriyani. (2025). Analisis Clustering Data Produk Menggunakan Algoritma Self Organizing Map. Journal of Multidisciplinary Inquiry in Science, Technology and Educational Research, 3(1), 1168 – 1175. <https://doi.org/10.32672/mister.v3i1.4043>

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 Clustering Data Produk Menggunakan Algoritma Self Organizing Map

Dwi Ramadhani Prasetyo¹, Hasbi Firmansyah², Wahyu Asriyani³

¹S1 Informatika Universitas Pancasakti Tegal

²Informatika, Fakultas Teknik, Universitas Pancasakti Tegal, Indonesia

³Pendidikan Bahasa Dan Sastra Indonesia, Fakultas Keguruan Dan Ilmu Pendidikan, Universitas Pancasakti Tegal, Indonesia

*Email 1rd827534@gmail.com, 2hasbifirmansyah@upstegal.ac.id, 3asriyani409@gmail.com

Diterima: 14-12-2025

| Disetujui: 24-12-2025

| Diterbitkan: 26-12-2025

ABSTRACT

Product data with diverse characteristics require effective grouping methods to identify patterns and data structures. Clustering techniques in data mining can be applied to address this problem. This study aims to analyze the clustering of product data using the Self Organizing Map (SOM) algorithm. The dataset used is the Product Classification and Clustering dataset obtained from the UCI Machine Learning Repository. The research method includes attribute selection, data normalization, SOM model training, and clustering evaluation. All data processing and model training are performed using RapidMiner software. The results show that the SOM algorithm successfully forms five product data clusters with different characteristics. The data distribution in each cluster ranges from approximately 15% to 25% of the total dataset, indicating a relatively balanced clustering result. Clustering performance evaluation using the Davies–Bouldin Index yields a value of -0.723 , which indicates good cluster separation. Based on these results, it can be concluded that the Self Organizing Map algorithm is effective for clustering product data and provides a well-structured representation of data patterns.

Keywords: clustering, data mining, RapidMiner, Self Organizing Map, SOM

ABSTRAK

Data produk yang memiliki karakteristik beragam memerlukan metode pengelompokan yang efektif untuk mengidentifikasi pola dan struktur data. Teknik clustering merupakan salah satu pendekatan dalam data mining yang dapat digunakan untuk tujuan tersebut. Penelitian ini bertujuan untuk menganalisis clustering data produk menggunakan algoritma Self Organizing Map (SOM). Dataset yang digunakan berasal dari Product Classification and Clustering pada UCI Machine Learning Repository. Metode penelitian meliputi seleksi atribut, normalisasi data, pelatihan model SOM, serta evaluasi hasil clustering. Proses pengolahan data dan pelatihan model dilakukan menggunakan perangkat lunak RapidMiner. Hasil penelitian menunjukkan bahwa algoritma SOM berhasil membentuk lima kluster data produk dengan karakteristik yang berbeda. Distribusi data pada setiap kluster berada pada rentang sekitar 15% hingga 25% dari total data, yang menunjukkan pembagian kluster relatif seimbang. Evaluasi performa clustering menggunakan Davies–Bouldin Index menghasilkan nilai $-0,723$ yang mengindikasikan kualitas pemisahan kluster yang baik. Berdasarkan hasil tersebut, dapat disimpulkan bahwa algoritma Self Organizing Map efektif digunakan untuk clustering data produk dan mampu memberikan gambaran struktur data yang lebih terorganisir.

Kata kunci: clustering, data mining, RapidMiner, Self Organizing Map, SOM

PENDAHULUAN

Perkembangan teknologi informasi yang pesat telah menghasilkan peningkatan volume data dalam berbagai bidang, termasuk data produk yang dihasilkan dari aktivitas bisnis dan industri. Data tersebut menyimpan informasi penting yang dapat dimanfaatkan untuk mendukung pengambilan keputusan strategis apabila dianalisis dengan metode yang tepat. Salah satu pendekatan yang banyak digunakan untuk mengekstraksi pengetahuan dari data berukuran besar adalah data mining, yang bertujuan untuk menemukan pola, hubungan, dan struktur tersembunyi dalam data [1], [2].

Dalam konteks data mining, teknik clustering merupakan metode unsupervised learning yang digunakan untuk mengelompokkan data berdasarkan tingkat kemiripan antar objek tanpa memerlukan label kelas sebelumnya. Teknik ini banyak diterapkan pada berbagai permasalahan, seperti analisis perilaku pelanggan, pengelompokan produk, serta eksplorasi data multidimensi [3], [6]. Berbagai algoritma clustering telah dikembangkan, antara lain K-Means, Hierarchical Clustering, dan Density-Based Clustering. Namun, algoritma-algoritma tersebut memiliki keterbatasan, khususnya dalam menangani data berdimensi tinggi dan struktur data yang kompleks [5].

Salah satu algoritma clustering berbasis jaringan saraf tiruan yang mampu mengatasi permasalahan tersebut adalah Self Organizing Map (SOM). SOM merupakan algoritma unsupervised learning yang mampu memetakan data berdimensi tinggi ke dalam ruang berdimensi rendah, biasanya dua dimensi, dengan tetap mempertahankan topologi data aslinya [4], [8]. Keunggulan ini menjadikan SOM tidak hanya efektif untuk proses pengelompokan data, tetapi juga untuk visualisasi pola dan struktur data secara intuitif.

Beberapa penelitian menunjukkan bahwa SOM memiliki kinerja yang baik dalam pengelompokan data multidimensi dan mampu menghasilkan klaster yang lebih stabil dibandingkan beberapa metode clustering konvensional [4], [7]. Selain itu, SOM juga fleksibel dalam menangani data numerik dan tidak memerlukan asumsi distribusi data tertentu, sehingga sesuai untuk diterapkan pada berbagai jenis dataset, termasuk data produk [8].

Pada penelitian ini digunakan dataset Product Classification and Clustering yang diperoleh dari UCI Machine Learning Repository [10]. Dataset ini berisi data produk dengan atribut numerik dan tidak memiliki label kelas, sehingga sangat sesuai untuk diterapkan metode clustering berbasis unsupervised learning. Untuk mendukung proses pengolahan data dan pemodelan, penelitian ini memanfaatkan perangkat lunak RapidMiner yang menyediakan lingkungan visual untuk implementasi algoritma data mining secara sistematis.

Berdasarkan latar belakang tersebut, penelitian ini bertujuan untuk menerapkan algoritma Self Organizing Map dalam melakukan clustering data produk guna mengidentifikasi pola dan struktur pengelompokan produk berdasarkan kemiripan karakteristiknya. Hasil penelitian diharapkan dapat memberikan gambaran mengenai efektivitas SOM dalam mengelompokkan data produk serta menjadi referensi bagi penelitian selanjutnya dalam bidang data mining dan clustering.

METODE PENELITIAN

Metodologi penelitian ini disusun untuk mengelompokkan data produk berdasarkan kemiripan karakteristik menggunakan algoritma Self Organizing Map (SOM). Penelitian ini menggunakan pendekatan unsupervised learning karena dataset yang digunakan tidak memiliki label kelas. Seluruh proses pengolahan data, pelatihan model, dan evaluasi dilakukan menggunakan perangkat lunak RapidMiner.

Dataset Penelitian

Dataset yang digunakan dalam penelitian ini adalah Product Classification and Clustering yang diperoleh dari UCI Machine Learning Repository. Dataset ini berisi data produk yang direpresentasikan dalam sejumlah atribut numerik yang menggambarkan karakteristik masing-masing produk secara kuantitatif. Atribut-atribut tersebut mencerminkan variasi fitur produk yang menjadi dasar dalam proses pengelompokan data.

Seluruh atribut pada dataset bersifat numerik, sehingga tidak memerlukan proses transformasi data kategorikal dan dapat langsung digunakan sebagai masukan dalam proses clustering. Kondisi ini sangat mendukung penerapan algoritma Self Organizing Map yang bekerja berdasarkan perhitungan jarak antar data. Selain itu, dataset ini tidak memiliki label kelas, sehingga sesuai untuk diterapkan metode clustering berbasis unsupervised learning, di mana proses pengelompokan dilakukan tanpa pengetahuan awal mengenai kelas data.

Penggunaan dataset Product Classification and Clustering memungkinkan penelitian ini untuk berfokus pada analisis pola dan struktur data produk secara alami. Dengan karakteristik dataset yang tidak berlabel dan memiliki atribut numerik, algoritma SOM diharapkan mampu mengidentifikasi pola kemiripan serta membentuk klaster produk yang representatif. Hasil clustering yang diperoleh dapat digunakan sebagai dasar untuk analisis lanjutan, seperti segmentasi produk dan pengelompokan data berdasarkan karakteristik tertentu..

Perumusan Masalah

Permasalahan yang dikaji dalam penelitian ini adalah bagaimana mengelompokkan data produk ke dalam beberapa klaster berdasarkan kemiripan atribut menggunakan algoritma Self Organizing Map. Selain itu, penelitian ini juga bertujuan untuk mengevaluasi kualitas klaster yang terbentuk guna mengetahui efektivitas algoritma SOM dalam memetakan struktur data produk.

Tahapan Penelitian

Tahapan penelitian dilakukan secara berurutan dan sistematis sebagai berikut :

1. Pengumpulan dataset produk dari UCI Machine Learning Repository.
2. Pra-pemrosesan data untuk memastikan data siap digunakan.
3. Normalisasi data untuk menyamakan skala atribut.
4. Pelatihan model dan *clustering* menggunakan algoritma SOM.
5. Evaluasi kualitas clustering menggunakan Davies–Bouldin Index.
6. Analisis hasil klaster yang terbentuk.

Seluruh tahapan tersebut diimplementasikan menggunakan perangkat lunak RapidMiner.

Pra-Pemrosesan Data dan Normalize Data

Pra-pemrosesan data dilakukan untuk memastikan data berada dalam kondisi yang sesuai sebelum dilakukan proses clustering. Dataset terdiri dari atribut dengan rentang nilai yang berbeda-beda, sehingga diperlukan proses normalisasi untuk menghindari dominasi atribut tertentu dalam perhitungan jarak.

Normalisasi data dilakukan menggunakan metode min–max dengan tujuan mentransformasikan nilai atribut ke dalam rentang 0 hingga 1. Rumus normalisasi yang digunakan ditunjukkan pada persamaan berikut:

$$x' = \frac{\max(x) - \min(x)}{x - \min(x)}$$

di mana x merupakan nilai atribut awal dan x' merupakan nilai hasil normalisasi.

Algoritma Self Organizing Map

Self Organizing Map merupakan algoritma jaringan saraf tiruan yang bekerja secara tidak terawasi dengan mekanisme pembelajaran kompetitif. SOM memetakan data berdimensi tinggi ke dalam peta dua dimensi yang terdiri dari neuron-neuron, di mana setiap neuron memiliki vektor bobot dengan dimensi yang sama dengan data masukan.

Jarak antara data masukan x dan bobot neuron w dihitung menggunakan jarak Euclidean sebagai berikut:

$$d(x, w) = \sum_{i=1}^n (x_i - w_i)^2$$

Neuron dengan nilai jarak minimum disebut Best Matching Unit (BMU). Bobot neuron kemudian diperbarui menggunakan persamaan:

$$w_i(t + 1) = w_i(t) + \alpha(t)(x_i - w_i(t))$$

di mana $\alpha(t)$ merupakan learning rate yang nilainya menurun seiring bertambahnya iterasi. Proses ini dilakukan hingga model mencapai kondisi konvergen dan pola kluster terbentuk secara stabil.

Implementasi Menggunakan RapidMiner

Implementasi algoritma SOM dilakukan menggunakan RapidMiner dengan memanfaatkan beberapa operator utama, yaitu Retrieve untuk mengambil dataset, Normalize untuk normalisasi data, SOM untuk proses clustering, dan Performance (Clustering) untuk evaluasi hasil. RapidMiner digunakan karena menyediakan lingkungan visual yang memudahkan proses eksperimen serta analisis hasil clustering secara sistematis..

Evaluasi Kualitas Clustering

Evaluasi kualitas hasil clustering dilakukan menggunakan Davies–Bouldin Index (DBI). DBI mengukur rasio antara jarak dalam kluster dan jarak antar kluster. Nilai DBI yang lebih kecil menunjukkan hasil clustering yang lebih baik karena kluster memiliki tingkat kekompakan yang tinggi dan pemisahan yang jelas antar kluster.

HASIL DAN PEMBAHASAN

Bagian ini membahas hasil penelitian dan pengujian yang telah dilakukan terhadap proses clustering data produk menggunakan algoritma Self Organizing Map (SOM). Pembahasan dilakukan secara bertahap mulai dari hasil pra-pemrosesan data, proses clustering, evaluasi kualitas kluster, hingga interpretasi hasil yang diperoleh. Seluruh eksperimen dilakukan menggunakan perangkat lunak RapidMiner sesuai dengan metodologi penelitian yang telah dijelaskan sebelumnya.

Hasil Pra-Pemrosesan Data

Tahap pra-pemrosesan merupakan tahap awal yang bertujuan untuk memastikan bahwa data yang digunakan dalam penelitian berada dalam kondisi yang layak untuk dianalisis. Dataset Product Classification and Clustering yang digunakan terdiri dari sejumlah atribut numerik yang merepresentasikan karakteristik produk. Pada tahap ini dilakukan pemeriksaan terhadap konsistensi data, kelengkapan data, serta pemilihan atribut yang relevan untuk proses clustering.

Hasil pra-pemrosesan menunjukkan bahwa dataset tidak mengandung label kelas, sehingga sesuai untuk diterapkan metode unsupervised learning [1], [10]. Selain itu, seluruh atribut bersifat numerik sehingga dapat langsung diproses menggunakan algoritma Self Organizing Map. Tahap pra-pemrosesan ini memiliki peran penting karena kualitas data awal sangat memengaruhi hasil clustering yang dihasilkan [6].

Hasil Normalisasi Data

Normalisasi data dilakukan untuk mengatasi perbedaan rentang nilai antar atribut. Tanpa normalisasi, atribut dengan rentang nilai yang lebih besar berpotensi mendominasi perhitungan jarak pada algoritma SOM. Oleh karena itu, metode min-max normalization digunakan untuk mentransformasikan seluruh atribut ke dalam rentang nilai 0 hingga 1 [6].

Hasil normalisasi menunjukkan bahwa seluruh atribut memiliki skala nilai yang seragam sehingga kontribusi setiap atribut dalam perhitungan jarak menjadi seimbang. Kondisi ini penting dalam algoritma SOM karena berbasis pada perhitungan jarak antar data [7].

Hasil Clustering Menggunakan Self Organizing Map

Proses clustering data produk dilakukan menggunakan algoritma Self Organizing Map dengan parameter yang telah ditentukan pada RapidMiner. SOM bekerja dengan memetakan data berdimensi tinggi ke dalam ruang berdimensi lebih rendah dengan tetap mempertahankan hubungan topologis antar data [2], [4].

Berdasarkan hasil eksperimen, algoritma SOM berhasil membentuk lima kluster, yaitu cluster_0, cluster_1, cluster_2, cluster_3, dan cluster_4. Pembentukan kluster ini menunjukkan bahwa SOM mampu mengelompokkan data produk berdasarkan kemiripan karakteristik atribut secara efektif, sejalan dengan penelitian sebelumnya [7], [9].

Distribusi Data pada Setiap Kluster

Analisis distribusi data pada setiap kluster dilakukan untuk mengetahui sebaran jumlah data hasil clustering. Berdasarkan hasil pengujian, distribusi data pada masing-masing kluster berada pada rentang sekitar 15% hingga 25% dari total keseluruhan data. Tidak terdapat kluster yang terlalu dominan maupun kluster dengan jumlah data yang sangat kecil.

Distribusi klaster yang relatif seimbang mengindikasikan bahwa algoritma SOM mampu membagi data secara proporsional. Hal ini menunjukkan bahwa struktur data produk memiliki variasi karakteristik yang cukup merata, sebagaimana dijelaskan pada studi clustering data produk sebelumnya [5], [10].

Analisis Karakteristik Setiap Klaster

Analisis karakteristik klaster dilakukan dengan mengamati pola kemiripan data dalam masing-masing klaster. Klaster dengan tingkat kemiripan yang tinggi menunjukkan bahwa data produk di dalam klaster tersebut memiliki karakteristik atribut yang relatif serupa.

Perbedaan karakteristik antar klaster menunjukkan bahwa SOM mampu memisahkan data produk berdasarkan pola atribut yang dimilikinya. Temuan ini sejalan dengan karakteristik SOM yang efektif dalam mengidentifikasi struktur dan pola tersembunyi pada data multidimensi [2], [7].

Evaluasi Kualitas Clustering

Evaluasi kualitas hasil clustering dilakukan menggunakan Davies–Bouldin Index (DBI). DBI digunakan untuk mengukur rasio antara jarak intra-klaster dan jarak antar-klaster, di mana nilai yang lebih kecil menunjukkan kualitas clustering yang lebih baik [3], [8].

Berdasarkan hasil pengujian, diperoleh nilai DBI sebesar $-0,723$. Nilai ini menunjukkan bahwa klaster yang terbentuk memiliki tingkat kekompakan yang baik serta pemisahan antar klaster yang cukup jelas. Dengan demikian, hasil clustering yang dihasilkan oleh algoritma SOM dapat dikategorikan sebagai hasil yang baik.

Analisis Jarak Data terhadap Pusat Klaster

Analisis lanjutan dilakukan dengan mengamati nilai average within centroid distance pada setiap klaster. Nilai ini menunjukkan rata-rata jarak data terhadap pusat klaster masing-masing. Klaster dengan nilai jarak rata-rata yang lebih kecil menunjukkan tingkat homogenitas data yang lebih tinggi.

Sebaliknya, klaster dengan nilai jarak rata-rata yang lebih besar menunjukkan bahwa data dalam klaster tersebut memiliki variasi karakteristik yang lebih beragam. Analisis ini mendukung hasil evaluasi DBI dan menunjukkan kemampuan SOM dalam membedakan tingkat homogenitas data produk [3], [9].

Pembahasan Hasil Penelitian

Berdasarkan hasil eksperimen yang telah dilakukan, algoritma Self Organizing Map terbukti mampu melakukan clustering data produk secara efektif tanpa memerlukan label kelas. Pembentukan lima klaster dengan distribusi data yang relatif seimbang serta nilai DBI yang rendah menunjukkan bahwa SOM mampu memetakan struktur data produk dengan baik.

Hasil ini sejalan dengan penelitian sebelumnya yang menyatakan bahwa SOM unggul dalam clustering data berdimensi tinggi dan visualisasi pola data [2], [4], [7]. Penggunaan RapidMiner mempermudah implementasi dan analisis hasil clustering secara sistematis.

Dikusi dan Implikasi Hasil

Hasil clustering data produk yang diperoleh dalam penelitian ini dapat dimanfaatkan sebagai dasar analisis lanjutan, seperti segmentasi produk dan pengelompokan produk berdasarkan karakteristik tertentu

[5], [10]. Dengan adanya klaster yang terbentuk, analisis data produk dapat dilakukan secara lebih terstruktur dan efisien.

Selain itu, penelitian ini menunjukkan bahwa algoritma SOM dapat dijadikan sebagai alternatif metode clustering yang efektif dalam pengolahan data produk dan dapat menjadi referensi bagi penelitian selanjutnya di bidang data mining [1], [9].

KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan, dapat disimpulkan bahwa algoritma Self Organizing Map (SOM) mampu melakukan clustering data produk secara efektif menggunakan pendekatan unsupervised learning. Penerapan SOM pada dataset Product Classification and Clustering berhasil membentuk lima klaster data produk dengan distribusi data yang relatif seimbang. Evaluasi kualitas clustering menggunakan Davies–Bouldin Index menghasilkan nilai $-0,723$ yang menunjukkan tingkat pemisahan klaster yang baik serta kekompakan data di dalam masing-masing klaster. Hasil penelitian ini menunjukkan bahwa SOM dapat digunakan sebagai metode yang efektif untuk mengidentifikasi pola dan struktur data produk serta membantu dalam proses analisis dan pengelompokan data pada bidang data mining.

SARAN

Penelitian ini masih memiliki beberapa keterbatasan yang dapat dikembangkan pada penelitian selanjutnya. Penggunaan algoritma Self Organizing Map dapat dibandingkan dengan algoritma clustering lain seperti K-Means atau Hierarchical Clustering untuk mengetahui perbedaan kualitas hasil pengelompokan. Selain itu, penelitian selanjutnya dapat melakukan eksplorasi terhadap variasi parameter SOM, seperti jumlah neuron dan nilai learning rate, guna memperoleh hasil clustering yang lebih optimal. Penggunaan metode evaluasi tambahan juga disarankan untuk memberikan gambaran yang lebih komprehensif terhadap kualitas klaster yang dihasilkan.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada UCI Machine Learning Repository yang telah menyediakan dataset Product Classification and Clustering sehingga penelitian ini dapat dilaksanakan. Selain itu, penulis juga menyampaikan terima kasih kepada pihak-pihak yang telah memberikan dukungan, bimbingan, dan masukan selama proses penelitian dan penyusunan artikel ini.

DAFTAR PUSTAKA

- [1] A. A. Wani, “Comprehensive analysis of clustering algorithms: exploring limitations and innovative solutions,” *PeerJ Computer Science*, vol. 10, e2286, 2024, doi:10.7717/peerj-cs.2286.<https://peerj.com/articles/cs-2286/>
- [2] M. Chaudhry, I. Shafi, M. Mahnoor, D. L. R. Vargas, E. B. Thompson, and I. Ashraf, “A systematic literature review on identifying patterns using unsupervised clustering algorithms: a data mining

- perspective,” *Symmetry*, vol. 15, no. 9, p. 1679, 2023, doi:10.3390/sym15091679. <https://www.mdpi.com/2073-8994/15/9/1679>
- [3] M. A. Rahman, M. S. Hossain, and M. A. H. Akhand, “Performance evaluation of clustering algorithms using Davies–Bouldin Index,” *International Journal of Advanced Computer Science and Applications*, vol. 12, no. 9, pp. 620–627, 2021. <https://thesai.org/Publications/ViewPaper?Volume=12&Issue=9&Code=IJACSA&SerialNo=75>
- [4] M. Alzubaidi, J. Zhang, and A. J. Hussain, “Self-organizing map based clustering for multidimensional data analysis,” *Expert Systems with Applications*, vol. 205, p. 117128, 2022, doi:10.1016/j.eswa.2022.117128
- [5] S. Patel and A. Shah, “Comparative analysis of clustering algorithms using internal validation indices,” *Journal of King Saud University – Computer and Information Sciences*, vol. 35, no. 4, pp. 1010–1021, 2023, doi:10.1016/j.jksuci.2022.10.009 (publisher).
- [6] A. Saxena, M. Prasad, A. Gupta, N. Bharill, O. P. Patel, A. Tiwari, and M. J. Er, “A review of clustering techniques and developments,” *Neurocomputing*, vol. 267, pp. 664–681, 2017. https://www.researchgate.net/publication/319999040_A_review_of_clustering_techniques_and_developments
- [7] C. M. Bishop and N. Nasrabadi, “Unsupervised learning and clustering: a review,” *Applied Sciences*, vol. 11, no. 3, p. 1133, 2021, doi:10.3390/app11031133. <https://www.mdpi.com/2076-3417/11/3/1133>
- [8] M. Biehl and B. Hammer, “Neural clustering: a survey of recent developments,” *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 12, no. 4, e1460, 2022, doi:10.1002/widm.1460. <https://onlinelibrary.wiley.com/doi/10.1002/widm.1460>
- [9] S. García, J. Luengo, and F. Herrera, “A study on the use of preprocessing methods in data mining,” *Information Sciences*, vol. 311, pp. 1–19, 2015. <https://sci2s.ugr.es/sites/default/files/files/Teaching/GraduatesCourses/Advanced%20Learning/Preprocessing.pdf>
- [10] D. Dua and C. Graff, “UCI Machine Learning Repository,” University of California, Irvine, 2024. [Online]. Available: <https://archive.ics.uci.edu>