

Penerapan *Clustering* Menggunakan Metode K-Means Untuk Penggunaan *E-Learning* Di Dunia

Marshanda Amalia Vega^{1*}, Via Kris Savitri¹, Terttiaavini²

¹Fakultas Ilmu Komputer, Teknik Informatika, Universitas Indo Global Mandiri, Palembang, Indonesia

²Fakultas Ilmu Komputer, Sistem Informasi, Universitas Indo Global Mandiri, Palembang, Indonesia

Email: ^{1*}2021110097@students.uigm.ac.id, ^{2*}2021110020@students.uigm.ac.id,

³avini.saputra@uigm.ac.id

(* : coresponding author)

Abstrak– Penelitian ini menjelaskan tentang penerapan metode K-means Clustering untuk menganalisis data pengguna e-learning. E-learning adalah bentuk pembelajaran yang menggunakan media berbasis elektronik. Tujuan utama dari penelitian ini adalah untuk mengelompokkan pengguna e-learning berdasarkan kesamaan atribut tertentu dan menemukan pola pada data. Langkah-langkah penelitian meliputi pengumpulan data pengguna e-learning dari website keaglee, dari Januari 2004 hingga Oktober 2021, pembersihan data untuk memastikan keakuratan dan konsistensi, dan penerapan Algoritma clustering. Algoritma ini membagi data kedalam beberapa kelompok berdasarkan kesamaan. Dalam penelitian ini, data dibagi menjadi tiga kelompok menggunakan nilai $k=3$ Melalui pengujian dengan metode davies bouldin, ditemukan hasil terbaik diperoleh pada cluster ke-9 dengan centroid 1.279. Cluster ini memiliki karakteristik pengguna e-learning yang serupa. metode K-means Clustering berhasil menganalisis data pengguna e-learning dengan sederhana, efisien, dan mudah diinterpretasikan. Pengelompokkan pengguna e-learning berdasarkan atribut serupa dapat dilakukan menggunakan metode ini. Penelitian ini dapat menjadi dasar untuk pengembangan lebih lanjut dalam penggunaan metode clustering dalam e-learning.

Kata Kunci: K-Means, *E-Learning*, *Clustering*, Algoritma

Abstract– This research describes the application of the K-means Clustering method to analyze e-learning user data. E-learning is a form of learning that uses electronic-based media. The main objective of this research is to cluster e-learning users based on the similarity of certain attributes and find patterns in the data. The research steps include collecting e-learning user data from keaglee website, from January 2004 to October 2021, cleaning the data to ensure accuracy and consistency, and applying clustering algorithm. This algorithm divides data into groups based on similarities. In this study, the data was divided into three groups using a value of $k = 3$. Through testing with the davies bouldin method, the best results were found in the 9th cluster with a centroid of 1,279. This cluster has similar e-learning user characteristics. K-means Clustering method successfully analyzes e-learning user data simply, efficiently, and easily interpreted. Grouping e-learning users based on similar attributes can be done using this method. This research can be the basis for further development in the use of clustering methods in e-learning

Keywords: K-Means, *E-Learning*, *Clustering*, Algoritma

1. PENDAHULUAN

Pendidikan jarak jauh merupakan suatu jenis pendidikan dimana kegiatan belajar mengajar antara pengajar dan peserta dilakukan di tempat yang berbeda, yang menggunakan sarana untuk mendukung proses belajar mengajar. Dengan demikian pendidikan jarak jauh disebut sebagai media e-learning yang memiliki persamaan istilah. Di antaranya adalah online learning, networked learning, virtual learning, dan distance learning (pembelajaran jarak jauh) (Sari, 2015).

Perkembangan penggunaan e-learning di seluruh dunia telah mengubah cara kita mengakses dan mengambil bagian dalam proses pendidikan. Dengan kemajuan teknologi informasi dan komunikasi, e-learning telah menjadi pilihan yang sangat populer di kalangan individu, lembaga pendidikan, dan organisasi dari berbagai sektor. E-learning memberikan fleksibilitas dan kebebasan dalam belajar, memungkinkan akses ke pengetahuan dan keterampilan tanpa adanya batasan geografis atau pembatasan waktu. Kemajuan teknologi mendorong adanya inisiatif untuk mengadopsi hasil-hasil teknologi dalam proses pendidikan. Para guru diharapkan memiliki kemampuan untuk memanfaatkan dan mengembangkan berbagai media pembelajaran yang tersedia (Saputra, 2017).

E-learning adalah sebuah proses pemanfaatan media berbasis elektronik untuk kegiatan belajar mengajar menggunakan jaringan Komputer. Dengan dikembangkannya jaringan komputer memungkinkan untuk dikembangkannya juga proses belajar mengajar berbasis web, dapat dikembangkan ke dalam jaringan Komputer (Sugiono, 2019).

Namun, dengan pertumbuhan pesat jumlah siswa dan ketersediaan beragam materi pembelajaran dalam platform e-learning, mengelola dan menyajikan konten yang relevan dan disesuaikan telah menjadi tantangan yang semakin kompleks. Untuk mengatasi masalah ini, diperlukan teknik-teknik analisis data yang efisien dan efektif.

Salah satu teknik analisis data yang digunakan dalam e-learning adalah clustering. Clustering adalah suatu metode yang mengelompokkan data ke dalam kelompok-kelompok yang serupa berdasarkan atribut atau karakteristik yang mirip. Salah satu metode clustering yang terkenal dan sering digunakan adalah metode k-means. Dalam jurnal ini kami melakukan eksplorasi data dengan menggunakan data mining. Tujuan dari Data mining adalah untuk menggali data dengan menggunakan metode yang sesuai dengan tujuan (Tedy, 2020). penelitian ini juga akan meneliti penggunaan metode k-means clustering dalam konteks e-learning. Dalam konteks ini, kita akan melihat bagaimana metode ini dapat digunakan untuk menganalisis data pengguna e-learning, mengelompokkan pengguna menjadi kelompok-kelompok berdasarkan jumlah pengguna yang ada di Dunia. Dengan menerapkan clustering menggunakan metode k-means dalam e-learning.

2. METODOLOGI PENELITIAN

2.1 Dataset

Dalam pengumpulan data terdapat dua jenis sumber data, yaitu Data Primer dan sekunder. Data primer merujuk pada data yang diperoleh secara langsung dari objek penelitian, sedangkan Data Sekunder adalah data yang merujuk dari literature, buku referensi, maupun browsing internet (Andi Prastowo, 2012). Dan sumber data yang digunakan dalam penelitian ini disebut data sekunder, karena data ini diperoleh dari Kaggle.com dengan halaman web <https://www.kaggle.com/datasets/marlonferrari/elearning-student-reactions> (Novianti, 2019).

2.2 Data Cleaning

Data cleaning atau yang biasa dikenal dengan pembersihan data, merupakan proses mengoreksi atau menghapus kesalahan, ketidakakuratan, dan ketidak konsistenan dalam data. Tujuan dari data cleaning adalah untuk memastikan bahwa data yang digunakan dalam analisis atau pembelajaran mesin adalah akurat, konsisten, dan dapat dipercaya. Dengan melakukan pembersihan data secara teliti, kita dapat memastikan bahwa data yang digunakan memiliki kualitas yang tinggi dan dapat diandalkan dalam pengambilan keputusan yang tepat. dalam penelitian ini, kami melakukan data cleaning dengan menghapus data yang tidak digunakan.

2.3 Algoritma K-Means

K-Means adalah salah satu metode non-hierarkis dalam *clustering* yang bertujuan untuk membagi data menjadi satu atau lebih kelompok. Metode ini mengelompokkan data berdasarkan karakteristik yang sama dalam satu kelompok, dan data yang dengan karakteristik yang berbeda di kelompokkan ke dalam kelompok yang beda [6]. Algoritma *K-means* adalah sebuah algoritma pengelompokan (*clustering*) yang melakukan proses iterasi berulang hingga mencapai konvergensi. Pada setiap iterasi, algoritma ini menghitung pusat *cluster* (*centroid*) dan jarak antara setiap data dengan *centroid*. Data-data tersebut kemudian diklasifikasikan berdasarkan kedekatannya dengan *centroid* (Zamzam, 2018).

3. ANALISA DAN PEMBAHASAN

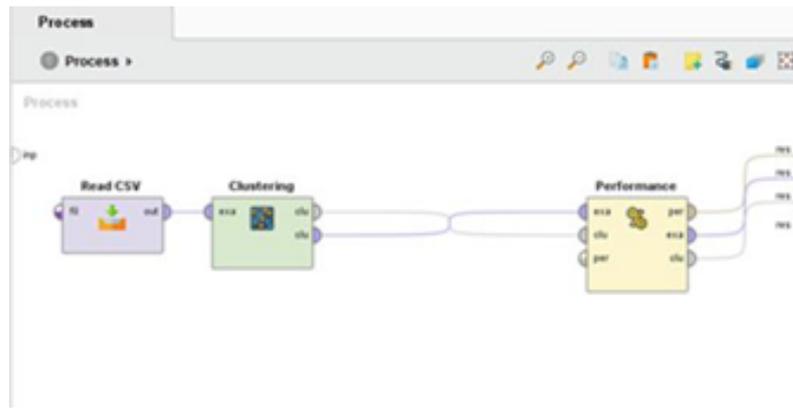
3.1 Pemilihan Model

Kami memilih menggunakan Algoritma *K-Means* karena merupakan salah satu klusterisasi sederhana dan efisien. Algoritma mudah dipahami dan diimplementasikan, serta cocok untuk data dengan jumlah atribut yang besar. Hasil dari Algoritma *K-Means* relative mudah diimplementasi-

kan, dan centroid dari setiap cluster dapat memberikan pemahaman yang jelas tentang karakteristik cluster tersebut. Algoritma K-Means juga dapat mengatasi data dengan berbagai jenis fitur. Baik numerik maupun kategorikal, meskipun secara umum lebih cocok untuk fitur numerik.

3.2 Penerapan Model

Dalam analisis data ini, kami menggunakan Algoritma *K-Means* untuk mengelompokkan data penggunaan *e-learning*. Setelah melakukan *preprocessing* data, kami menerapkan algoritma *K-Means* dengan menginputkan data dan mencari cluster menggunakan metode *K-Means*. Selanjutnya dengan cara menghubungkan hasil clustering dengan file CSV yang berisi data tersebut. Setelah itu, dilakukan evaluasi kinerja dan menghubungkan hasil *K-Means* dengan evaluasi kinerja tersebut agar dapat menjalankan proses clusterisasi pada aplikasi *Rapid Miner*.



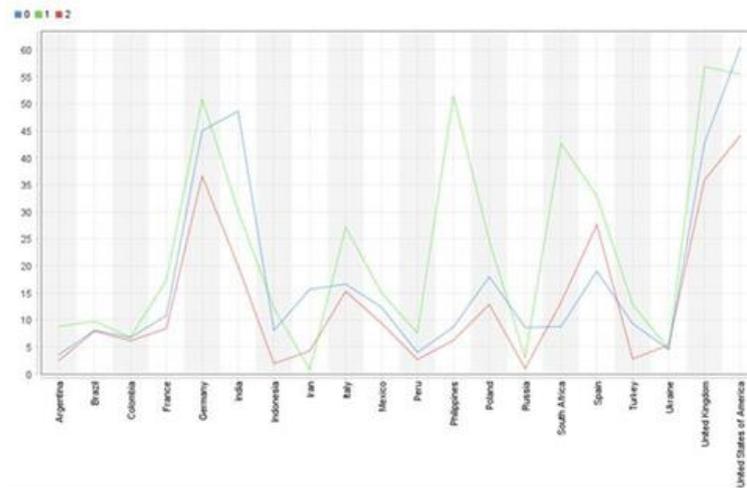
Gambar 1. Penerapan Model

Pilih nilai K yang akan digunakan untuk membagi data menjadi kelompok cluster. Dalam penelitian ini dipilih nilai $k=3$ untuk melakukan pengclusteran. Setelah itu, jalankan proses untuk mendapatkan hasil pengclusteran. Hasil cluster dapat dilihat pada centroid yang di tampilkan di bawah ini.

Tabel 1. Hasil *Clustering*

Atributte	Cluster_0	Cluster_1	Cluster_2
Argentina	35.344.827.586.206.800	8.785.714.285.714.280	24.453.125
Brazil	8.068.965.517.241.370	9.75	79.453.125
Colombia	6.724.137.931.034.480	6.785.714.285.714.280	61.328.125
France	10.775.862.068.965.500	17.25	8.390.625
Germany	4.498.275.862.068.960	50.857.142.857.142.800	36.578.125
India	4.858.620.689.655.170	30.142.857.142.857.100	19.984.375
Indonesia	8.051.724.137.931.030	12.178.571.428.571.400	18.984.375
Iran	15.655.172.413.793.100	0.7857142857142857	4.25
Italy	16.603.448.275.862.000	27.107.142.857.142.800	152.578.125
Mexico	12.310.344.827.586.200	15.0	9.375
Peru	39.827.586.206.896.500	7.535.714.285.714.280	265.625
Philippines	8.655.172.413.793.100	5.142.857.142.857.140	61.875
Poland	17.982.758.620.689.600	24.678.571.428.571.400	12.859.375
Russia	860.344.827.586.207	29.642.857.142.857.100	0.9921875
South Africa	8.724.137.931.034.480	4.267.857.142.857.140	132.265.625
Spain	1.903.448.275.862.060	32.857.142.857.142.800	27.578.125
Turkey	9.293.103.448.275.860	12.785.714.285.714.200	2.75
Ukraine	4.517.241.379.310.340	4.607.142.857.142.850	53.828.125
United Kingdom	42.758.620.689.655.100	5.682.142.857.142.850	358.671.875
United States of America	6.043.103.448.275.860	5.542.857.142.857.140	440.703.125

Selain dalam format table hasil pengclustering juga dapat divisualisasikan melalui berbagai bentuk visualisasi, salah satunya dapat dilihat pada gambar 2.



Gambar 2. Hasil *Clustering* Dalam Visualisasi Plot

Pada pengujian di atas terlihat bahwa cluster 0 memiliki 58 data yang disimbolkan dengan wana biru, *cluster* 1 memiliki 28 data yang disimbolkan dengan warna hijau, dan cluster 2 memiliki 128 data yang disimbolkan dengan warna merah.

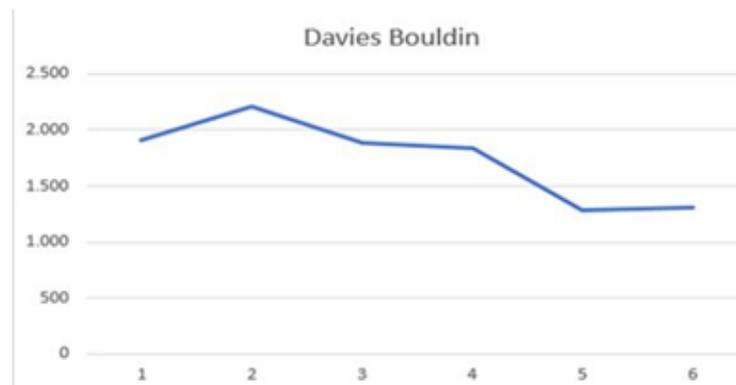
3.3 Pengujian

Pengujian ini menggunakan metode *Davies Bouldin* untuk mengukur jarak antara cluster yang dihasilkan dalam proses *clustering*. Pengukuran tersebut dilakukan dengan menghitung jarak antar centroid setiap cluster lainnya menggunakan aplikasi *Rapid Miner*.

Tabel 2. Cluster *Davies Bouldin*

K	Day
2	1.913
3	2.211
5	1.880
7	1.836
9	1.279
131.303	

Dengan menggunakan metode *Davies Bouldin* dan bantuan *Microsoft excel*, hasil pengujian menunjukkan bahwa cluster ke-9 memiliki nilai terbaik yaitu 1.279. informasi tersebut dapat dilihat ada grafik yang terlampir dalm gambar 3.



Gambar 3. Grafik *Davies Bouldin*

4. KESIMPULAN

Metode analisis yang digunakan adalah algoritma *K-means Clustering*. Yang bertujuan untuk mengelompokkan pengguna *e-learning* berdasarkan atribut yang serupa. Hasil dari metode ini memberikan pemahaman yang jelas tentang karakteristik kelompok pengguna.

Dalam hasil dan pembahasan, dipilih model *K-Means* sebagai metode analisis yang tepat karena sederhana, efisien, mudah diinterpretasikan, dan cocok untuk data dengan jumlah atribut yang besar. Model ini diaplikasikan pada data pengguna *e-learning* dengan $k=3$, membagi data menjadi tiga kelompok. Melalui pengujian, dengan menggunakan metode Davies Bouldin, hasil terbaik didapatkan pada *cluster* ke-9 dengan nilai centroid 1.303. *cluster* ini, dapat dilakukan analisis lebih lanjut untuk memahami karakteristik pengguna *e-learning* yang termasuk dalam kelompok tersebut.

Secara keseluruhan, penelitian ini berhasil menerapkan metode *K-means Clustering* dalam analisis data pengguna *e-learning*. Hasilnya memungkinkan pengelompokkan pengguna berdasarkan kesamaan atribut dan memberikan pemahaman yang lebih jelas tentang karakteristik setiap kelompok. Penelitian ini dapat menjadi dasar untuk pengembangan lebih lanjut dalam penggunaan metode *clustering* dalam konteks *e-learning*.

REFERENCES

- F. Sari, T. Avini, P. S. Informasi, P. S. Informasi, P. S. Informasi, and S. Informasi, (2015). "Sistem Informasi *e-Learning* pada SMK 6 Palembang Berbasis Web," vol. 1, pp. 21–30.
- T. S. Saputra and A. Fitriani, (2017). "PENINGKATAN KOMPETENSI GURU DALAM PEMBUATAN BAHAN AJAR MENGGUNAKAN MEDIA PEMBELAJARAN SMART LEARNING," vol. 1, no. 1, pp. 46–51.
- Sugiono, S. Nurdiani, S. Linawati, R. A. Safitri, and E. P. Saputra, (2019). "Pengelompokan Perilaku Mahasiswa Pada Perkuliahan *E-Learning* dengan *K-Means Clustering*," *J. Kaji. Ilm.*, vol. 19, no. 2, pp. 126–133.
- T. Tedy and S. Saputra, (2020). "Analisa Akurasi Penggunaan Metode Single Eksponensial Smoothing untuk Perkiraan Penerimaan Mahasiswa Baru Pada Perguruan Tinggi XYZ," vol. 11, no. 01, pp. 64–68.
- D. Novianti, (2019). "Implementasi Algoritma Naïve Bayes Pada Data Set Hepatitis Menggunakan Rapid Miner," vol. XXI, no. 1, doi: 10.31294/p.v20i2.
- D. Retno Lestari, (2019). "Implementasi Data Mining Untuk Pengaturan Layout Swalayan Delimas Lestari Kencana Dengan Menggunakan Dengan Menggunakan Algoritma *K-Means Clustering*," *Jurikom*, vol. 6, no. 2, pp. 112–117, [Online]. Available: <http://ejurnal.stmik-budidarma.ac.id/index.php/jurikom%7CPage%7C112>
- F. Zamzam, M. Ramadhan, A. K. Rosni, and T. S. Saputra, (2018). "Jurnal Internasional Teknik & Teknologi Analisis Pengelompokan Bidang Penelitian Utama," vol. 7, pp. 43–46.