

Clustering Struktur Pengetahuan pada Open-Ended Concept Map Menggunakan Algoritma K-Means

Agung Andhika Firdiansyah¹, Muhammad Iqbal Akbar², Didik Dwi Prasetya^{3*},

I Putu Arda Mahendra⁴

^{1,2,3,4}Universitas Negeri Malang

email: didikdwi@um.ac.id

Abstract: *Concept maps are graphical tools that can represent knowledge. A collection of several connected propositions will become a concept map where these propositions have a concept - link - concept structure. To be able to find out the structure of knowledge that is often discussed by students, clustering is carried out in this open-ended concept map. The clustering process uses the k-means algorithm which will divide the knowledge structure created by students and the dataset used is obtained from relational database material. Before processing the data to get the results of the knowledge structure that is often discussed, several stages are needed, namely data selection, data cleansing, data integration, data transformation, mining process, and evaluation. The data used has only 1 parameter, namely the proposition that has been processed from the concept table and the link table. Evaluation is used to measure the performance of the k-means clustering algorithm in clustering knowledge structure data on an open-ended concept map that has been made by students. From the data that has been clustered manually, it is identified with the results of clustering using the k-means algorithm to get performance results from clustering performance. The results obtained show that the k-means algorithm can cluster the knowledge structure in the open-ended concept map well, which has an accuracy of 86.13%.*

Keywords: *clustering, knowledge, open-ended concept map, k-means*

Abstrak: Peta konsep adalah alat grafis yang dapat mewakili representasi pengetahuan individu. Kumpulan dari beberapa proposisi yang terhubung akan menjadi sebuah peta konsep dimana proposisi-proposisi tersebut memiliki struktur konsep – link – konsep. Untuk dapat mengetahui struktur pengetahuan yang didiskusikan oleh siswa diperlukan operasi pengelompokan atau clustering pada peta konsep open-ended. Proses clustering menggunakan algoritma k-means yang akan membagi struktur pengetahuan yang dibuat oleh mahasiswa dan dataset yang digunakan diperoleh dari materi database relasional. Sebelum mengolah data untuk mendapatkan hasil struktur pengetahuan yang sering dibahas, diperlukan beberapa tahapan, yaitu pemilihan data, pembersihan data, integrasi data, transformasi data, proses penambangan, dan evaluasi. Data yang digunakan hanya memiliki 1 parameter yaitu proposisi yang telah diolah dari tabel konsep dan tabel link. Evaluasi digunakan untuk mengukur kinerja algoritma clustering k-means dalam mengklusterisasi data struktur pengetahuan pada peta konsep open-ended yang telah dibuat oleh siswa. Dari data yang telah di cluster secara manual diidentifikasi dengan hasil clustering menggunakan algoritma k-means untuk mendapatkan hasil kinerja dari kinerja clustering. Hasil yang diperoleh menunjukkan bahwa algoritma k-means dapat mengelompokkan struktur pengetahuan pada peta konsep open-ended dengan tingkat akurasi sebesar 86,13%.

Kata kunci: *clustering, knowledge, open-ended concept map, k-means*

Pendahuluan

Concept map adalah alat grafis yang dapat merepresentasikan sebuah pengetahuan. Pada umumnya memuat konsep yang ditutup dengan lingkaran atau kotak dan hubungan antar konsep yang ditunjukkan oleh garis penghubung (de Ries et al, 2021). Proposisi adalah gabungan antara dua concept dan satu link yang menjadi sebuah pernyataan yang bermakna (Prasetya dkk, 2020). Kumpulan dari beberapa proposisi yang terhubung akan menjadi suatu concept map yang dapat dikategorikan menjadi dua yaitu open-ended style dan closed-ended style. Menurut (Prasetya dkk, 2020) open-ended concept map memungkinkan peserta didik untuk menggunakan concept dan link yang diinginkan pada concept

map mereka, memberikan kebebasan untuk peserta didik dalam menambah, mengubah, maupun menghapus concept maupun link untuk merepresentasikan pengetahuan mereka.

Pada open-ended concept map, pengguna bebas menggunakan konsep maupun kata penghubung apa saja yang di dalam concept map, berbeda dengan close-ended concept map yang hanya diperbolehkan menggunakan konsep dan kata penghubung yang sudah ditentukan sebelumnya. (Hirashima, 2019). Menurut (de Ries et al., 2021) open-ended concept map pada bidang studi pendidikan, sosial maupun organisasi digunakan sebagai instrumen untuk mengoleksi data dan menganalisa pengetahuan konseptual dari individu. Menurut (Fatawi dkk, 2020) concept map ini merupakan strategi pengetahuan yang efektif dalam membantu mengorganisasikan konsep-konsep penting yang terkait dengan inti dari sebuah isu dan hubungannya antara konsep-konsep lainnya.

Pada open-ended concept map lebih sulit untuk dilakukan penilaian dan memberikan ulas balik karena akan terdapat banyak sekali variasi dari concept map yang dibuat. Oleh karena itu, diperlukan langkah-langkah praktis agar mampu mengungkap pola-pola representasi pengetahuan yang terbentuk. Tulisan ini tertarik untuk melakukan clustering struktur pengetahuan dari open-ended concept map guna melihat topik apa yang banyak dibahas oleh peserta didik dalam membuat concept map sehingga memudahkan guru menangkap pengetahuan pengguna.

Dalam penelitian ini akan dilakukan pengelompokan terhadap data struktur pengetahuan pada open-ended concept map. Clustering merupakan proses dalam menemukan pengetahuan dalam sekelompok data. Proses pengelompokan adalah pada data yang tidak berlabel, oleh karena itu, proses clustering tidak digunakan untuk memprediksi sebuah target (Iskandar, 2018). Clustering data akan mengelompokkan sejumlah data yang memiliki karakteristik yang sama (Aditya dkk, 2020). Tujuan dari clustering adalah untuk meminimalkan varietas dalam satu cluster dan memaksimalkan varietas antara satu cluster dengan cluster yang lainnya.

Proses clustering akan dilakukan menggunakan algoritma k-means. Algoritma k-means merupakan algoritma yang sederhana dan mudah dalam penerapannya (Hudin, 2018). Data akan dibagi menjadi beberapa cluster berbasis jarak (Sangga, 2018). Algoritma ini mengelompokkan data berdasarkan jarak data dan pusat cluster atau biasa disebut dengan centroid sebagai hasil dari proses iteratif, dimulai dengan pemilihan data k secara acak (Darmi & Setiawan, 2016). K disini merupakan banyaknya cluster yang akan di bentuk. Kelebihan dari algoritma k-means yaitu mudah digunakan dan relatif cepat dalam implementasinya, algoritma ini juga umum digunakan dalam penerapan tugas data mining (Rini dkk., 2016). Algoritma k-means diharapkan dapat melakukan clustering data struktur pengetahuan dengan baik dan dapat memberikan informasi pola pengetahuan dalam data tersebut.

Fokus dari penelitian ini adalah untuk melakukan klasterisasi data struktur pengetahuan pada open-ended concept map menggunakan algoritma k-means agar dapat mengetahui pembagian cluster yang terbentuk sehingga dapat ditemukan pola dalam data tersebut. Sebelum melakukan clustering menggunakan algoritma k-means, data yang awalnya terdiri dari concept dan link akan dihubungkan menjadi sebuah proposisi. Setelah itu akan dilakukan preprocessing terlebih dahulu guna untuk menghilangkan data-data yang bermasalah, menyamakan format data yang sudah dibuat oleh peserta didik, dan merubah data tersebut menjadi token. Setelah data dilakukan preprocessing maka dapat dilanjutkan untuk dilakukan proses clustering data dengan algoritma k-means menggunakan bantuan tools rapidminer sekaligus melihat visualisasi dari hasil clustering. Data yang digunakan dalam penelitian ini merupakan data penelitian pembuatan concept map dengan basis open-ended style dan pada materi basis data relasional dari setiap peserta didik. Hasil dari penelitian ini diharapkan dapat bermanfaat untuk berbagai pihak dan juga sebagai referensi bagi penelitian sejenis untuk kedepannya

Metode

Data yang digunakan pada penelitian ini adalah open-ended concept map. Data mentah berupa tabel concept dengan total 869 data dan tabel link dengan total 789 data. Data tersebut merupakan data penelitian pada pembuatan concept map dari materi matakuliah Basis Data yang ditempuh oleh mahasiswa semester 3. Pada penelitian ini berfokus pada mid 65 atau materi dengan tema basis data relasional. Atribut yang digunakan yaitu concept dan link pada masing-masing tabel.

Pada penelitian ini menggunakan enam tahapan data mining yaitu data selection, data integration, data cleansing, data transformation, proses mining, dan evaluasi. Tahap data selection melakukan pemilihan data yang akan digunakan untuk penelitian. Penelitian ini berfokus pada struktur pengetahuan peserta didik sehingga atribut yang akan digunakan yaitu hanya concept dan link yang sudah dibuat. Tercatat sebanyak 332 data untuk concept dan 298 data untuk link. Kedua data tersebut akan di proses untuk diubah struktur datanya menjadi proposisi.

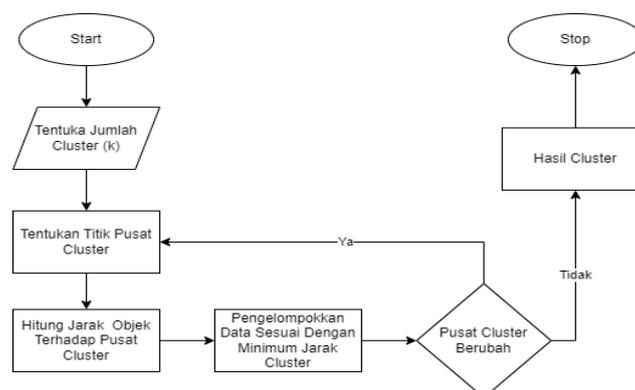
Selanjutnya, tahap data integration dilakukan untuk proses penggabungan data dari yang hanya concept dan link menjadi proposisi. Concept yang dihubungkan dengan concept lain dengan struktur concept – link – concept adalah yang disebut proposisi. Dalam proses penggabungan ini dibutuhkan juga atribut uid pada tabel concept beserta atribut source dan target pada tabel link. Link dengan source dan target akan dicocokkan dengan concept sehingga nanti akan membentuk suatu proposisi.

Tahap cleansing fokus pada aktivitas pembersihan data terhadap data proposisi yang bermasalah. Pembersihan data dilakukan untuk menghilangkan data kosong dan memperbaiki data yang terdapat kesalahan ejaan. Proses pembersihan data kosong dilakukan di phpMyAdmin menggunakan query mysql. Setelah data dibersihkan lalu dilanjutkan untuk dilakukan perbaikan data. Perbaikan data dilakukan di google docs menggunakan fitur spelling and grammar check. Hasil dari pembersihan dan perbaikan data yaitu data proposisi yang sudah tidak mengandung data kosong maupun kesalahan ejaan.

Pada tahap *transformation* dilakukan perubahan struktur data terhadap data proposisi yang akan dilakukan proses mining. Perubahan data dilakukan untuk menyamakan struktur data sehingga tidak ada perbedaan antara struktur masing-masing data yang menyebabkan ketidakakuratan pada proses mining. Proses transformasi ini dilakukan menggunakan bantuan aplikasi rapidminer dengan operator transform cases. Operator transform cases dapat mengubah teks menjadi uppercase maupun lowercase, tetapi dalam penelitian ini teks diubah menjadi lowercase

Clustering Open-Ended Concept Map

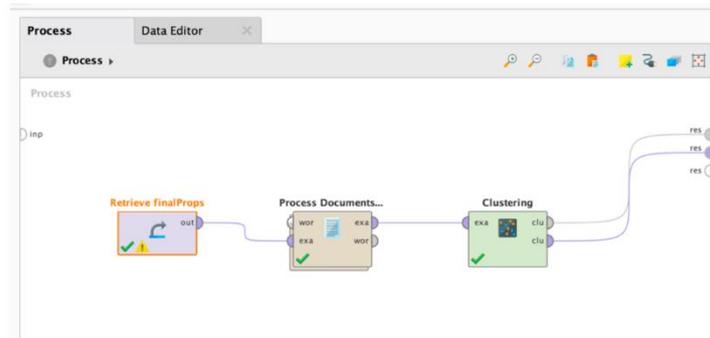
Pada tahap ini dilakukan clustering terhadap data yang sudah diproses sebelumnya, dan diilustrasikan pada Gambar 1. Clustering data menggunakan algoritma k-means dengan k sebanyak 5. Measure type yang digunakan pada penelitian ini yaitu numerical measure karena data sudah dilakukan text vectorization. Numerical measure hanya dapat digunakan untuk mengukur atribut dengan nilai numerik. Numerical measure yang digunakan yaitu cosine similarity yang akan menghitung kesamaan antara dua objek yang direpresentasikan dengan vector. Untuk mendapatkan hasil visual dari clustering yang dilakukan maka digunakan operator tambahan yaitu cluster model visualizer sehingga nanti akan dapat dilihat untuk hasil clustering dalam overview, heat map, centroid chart, centroid table, dan scatter plot.



Gambar 1. Tahapan clustering concept map

Tahapan yang pertama yaitu dengan memilih atau menentukan nilai k secara acak, lalu dilakukan penentuan titik pusat cluster. Setelah itu akan dilakukan perhitungan jarak antara data dengan titik pusat cluster dan dilakukan pengelompokan data sesuai dengan minimum jarak cluster. Lalu terdapat pengecekan jika pusat cluster berubah maka proses akan kembali ke tahap penentuan titik pusat cluster, tetapi jika tidak terdapat perubahan maka proses selesai dan didapatkan hasil clustering. Proses

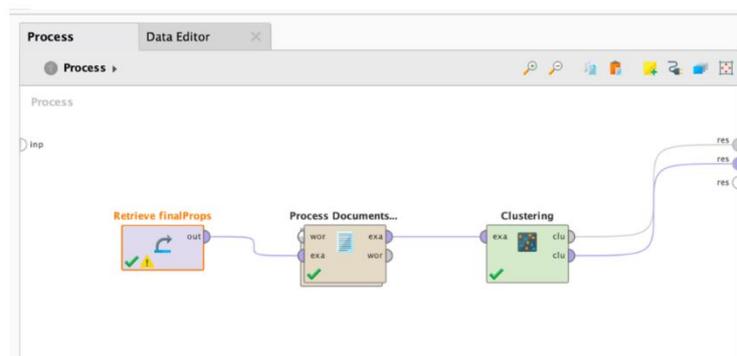
penentuan titik pusat cluster sampai dengan pengecekan perubahan titik pusat cluster dilakukan secara otomatis menggunakan algoritma k-means yang ada di rapidminer, seperti terlihat pada Gambar 2.



Gambar 2. Proses clustering concept map menggunakan RapidMiner

Hasil

Pada Gambar 3 menampilkan hasil dari clustering menggunakan algoritma k-means dengan k sebanyak 5 pada struktur pengetahuan open-ended concept map.



Gambar 3. Hasil clustering menggunakan RapidMiner

Berdasarkan kelompok cluster yang terbentuk, dapat diidentifikasi kandidat topik melalui dokumen concept map yang terlibat. Tabel 1 menunjukkan topik pembahasan apa saja dari masing-masing cluster yang didapatkan dari menyimpulkan keseluruhan item tiap cluster yang terbentuk.

Tabel 1. Kandidat topik cluster

No. cluster	Cluster candidate
2	Attribute/field/istilah
3	Foreign/super/alternate/candidate/primary
4	Istilah-istilah/relational key/pengertian
0	Keuntungan/sederhana/mudah
1	Kolom/baris/dimensi/cardinality

Pada Tabel 1 dapat dilihat bahwa telah didapatkan cluster candidate yang merupakan keyword sebagai acuan dalam melakukan clustering k-means. Cluster candidate pada tiap-tiap cluster dapat sama dengan cluster yang lain tetapi hal ini tidak berpengaruh terhadap hasil clustering karena data yang digunakan merupakan proposisi atau gabungan antara concept – link – concept sehingga meskipun ada beberapa kesamaan cluster candidate pada tiap cluster tetapi maksud dari proposisi pada tiap-tiap cluster dapat diklasterisasi dengan baik.

Pada tahap evaluasi dilakukan evaluasi terhadap kinerja dari clustering yang sudah dilakukan sebelumnya. Proses ini akan dilakukan perhitungan akurasi, presisi, dan recall menggunakan confusion matrix yang sudah dibuat dengan rumus standar yang telah dicantumkan.

Tabel 2. Confusion matrix hasil clustering

		Aktual	
		Positif	Negatif
Prediksi	Positif	212	18
	Negatif	20	24

Pada tabel 4.4 dihasilkan nilai True Positive (TP) sejumlah 212, True Negative (TN) sejumlah 24, False Positive (FP) sejumlah 18, dan False Negative (FN) sejumlah 20. Berdasarkan nilai yang sudah didapatkan diatas, dapat dilakukan perhitungan akurasi, presisi, dan recall dengan hasil sebagai berikut:

$$\begin{aligned} \text{Accuracy} &= 86.13\% \\ \text{Precision} &= 92.17\% \\ \text{Recall} &= 91.38\% \end{aligned}$$

Pembahasan

Setelah dilakukan perhitungan akurasi, presisi, dan recall untuk hasil kinerja clustering menggunakan algoritma k-means ini, didapatkan hasil akurasi sebesar 86,13%, presisi sebesar 92,17%, dan recall sebesar 91,38% yang cukup melampaui dari besaran akurasi. Hal tersebut disebabkan karena pada tabel confusion matrix terdapat perbandingan yang cukup jauh antara True Positive (TP) dengan True Negative (TN) sehingga juga berpengaruh pada hasil dari presisi maupun recall. Meskipun presisi dan recall melampaui akurasi, hasil dari clustering yang menggunakan k-means sudah cukup untuk membuktikan bahwa algoritma tersebut dapat dilakukan untuk melakukan clustering terhadap struktur pengetahuan. Hal tersebut terjadi karena hasil pada setiap cluster membahas suatu bahasan yang cukup relevan dan jumlah item yang tidak sesuai paling tinggi hanya 14 pada cluster 0.

Hasil akurasi membuktikan bahwa data proposisi pada open-ended concept map dapat dilakukan clustering menggunakan algoritma k-means karena mendapatkan akurasi yang cukup bagus. Hal ini disebabkan juga karena data yang siap pakai tersebut sudah melewati proses preprocessing sehingga meminimalisir untuk terdapat kecatatan data yang akan mempengaruhi dalam proses clustering yang dilakukan sehingga juga akan berpengaruh pada akurasinya.

Simpulan

Pada penelitian ini, dilakukan proses clustering terhadap struktur pengetahuan pada open-ended concept map menggunakan algoritma k-means dengan menggunakan parameter proposisi yang diperoleh dari preprocessing data yang awalnya concept dan link menjadi concept – link – concept. Proses ini dibantu dengan aplikasi rapidminer sehingga dapat dilakukan preprocessing dan proses mining data secara cepat dan dapat mengungkap ide-ide yang dibuat oleh siswa. Selanjutnya dilakukan evaluasi akurasi secara manual dari data proses mining yang sudah dihasilkan dari rapidminer. Didapatkan hasil rata-rata akurasi dari proses ini sebesar 86,13%.

Daftar Pustaka

- Aditya, A., Jovian, I., & Sari, B. N. (2020). Implementasi K-Means Clustering Ujian Nasional Sekolah Menengah Pertama di Indonesia Tahun 2018/2019. *Jurnal Media Informatika Budidarma*, 4(1), 51-58.
- Ageberg, E., Bunke, S., Lucander, K., Nilsen, P., & Donaldson, A. (2019). Facilitators to support the implementation of injury prevention training in youth handball: a concept mapping approach. *Scandinavian journal of medicine & science in sports*, 29(2), 275-285.
- Darmi, Y. D., & Setiawan, A. (2016). Penerapan metode clustering k-means dalam pengelompokan penjualan produk. *Jurnal Media Infotama*, 12(2).

- de Ries, K. E., Schaap, H., van Loon, A. M. M., Kral, M. M., & Meijer, P. C. (2021). A literature review of open-ended concept maps as a research instrument to study knowledge and learning. *Quality & Quantity*, 1-35.
- Eska, J. (2018). Penerapan Data Mining Untuk Prediksi Penjualan Wallpaper Menggunakan Algoritma C4. 5. *JURTEKSI (Jurnal Teknologi dan Sistem Informasi) Vol 2, Nomor 2, Maret 2016*, hlmn 9 – 13 ISSN 2407 - 1811
- Fatawi, I., Degeng, I. N. S., Setyosari, P., Ulfa, S., & Hirashima, T. (2020). Effect of online-based concept map on student engagement and learning outcome. *International Journal of Distance Education Technologies (IJDET)*, 18(3), 42-56.
- Fauzi, A. (2019, June). Penerapan Data Mining Dengan Teknik Clustering Menggunakan Algoritma K-Means Pada Data Transaksi Superst. In *SNIA (Seminar Nasional Informatika dan Aplikasinya) (Vol. 3, pp. D-15)*.
- Hirashima, T. (2019). Reconstructional concept map: automatic assessment and reciprocal reconstruction. *International Journal of Innovation, Creativity and Change*, 5, 669-682.
- Hudin, Muhammad Sholeh. "Implementasi Metode Text Mining dan K-Means Clustering untuk Pengelompokan Dokumen Skripsi (Studi Kasus: Universitas Brawijaya)." PhD diss., Universitas Brawijaya, 2018.
- Hutagalung, J. (2022). Pemetaan Siswa Kelas Unggulan Menggunakan Algoritma K-Means Clustering. *JATISI (Jurnal Teknik Informatika dan Sistem Informasi)*, 9(1), 606-620.
- Prasetya, D. D., Hirashima, T., & Hayashi, Y. (2020). Study on Extended Scratch- Build Concept Map to Enhance Students' Understanding and Promote Quality of Knowledge Structure. *International Journal of Advanced Computer Science and Applications*, 11(4).
- Prasetya, D., Hayashi, Y., & Hirashima, T. (2020). Improving Knowledge Structure through Extended Scratch-Build Concept Mapping. *Letters in Information Technology Education (LITE)*, 3(1), 36–40. <https://doi.org/10.17977/um010v3i12020p036>
- Pratama, E. E., & Sastypratiwi, H. (2021). Analisis kecenderungan informasi terkait COVID-10 berdasarkan big data sosial media dengan menggunakan metode data mining. *Jurnal Informatika Polinema*, 7(2), 1-6.
- Rini, F., & Kahar, N. (2016). Juliana, "Penerapan Algoritma K-Means Pada Pengelompokan Data Siswa Baru Berdasarkan Jurusan Di Smk Negeri 1 Kota Jambi Berbasis Web",. Semin. Nas. APTIKOM, 94-99.
- Sangga, V. A. P. (2018). Perbandingan Algoritma K-Means dan Algoritma K- Medoids dalam Pengelompokan Komoditas Peternakan di Provinsi
- Sudarsono, B. G., Leo, M. I., Santoso, A., & Hendrawan, F. (2021). Analisis Data Mining Data Netflix Menggunakan Aplikasi Rapid Miner. *JBASE-Journal of Business and Audit Information Systems*, 4(1).
- Wahyuni, W. A., & Saepudin, S. (2021, September). Penerapan Data Mining Clustering Untuk Mengelompokkan Berbagai Jenis Merk Mesin Cuci. In *Seminar Nasional Sistem Informasi dan Manajemen Informatika Universitas Nusa Putra (Vol. 1, No. 01, pp. 306-313)*.
- Zai, C. (2022). Implementasi Data Mining Sebagai Pengolahan Data. *Jurnal Portal Data*, 2(3).