Penerapan Algoritma *K-Means* Untuk *Clustering* Pemodelan Pengetahuan Pengguna Menggunakan Rapidminer

Application of K-Means Algorithm for Clustering User Knowladge Modeling Using Rapidminer

Khalid Alfa Yanuar[1], Hasbi Firmansyah[2]
[1,2] Universitas Pancasakti Tegal, Indonesia
Corresponding author : khalidalfay@gmail.com

Abstrak

Dalam penelitian ini, algoritma *K-Means* diterapkan untuk clustering data pengguna menggunakan rapidminer dalam rangka pemodelan pengetahuan. Lima variabel utama digunakan dalam dataset yang digunakan: waktu belajar untuk tujuan tertentu (STG), jumlah pengulangan untuk tujuan tertentu (SCG), waktu belajar untuk materi terkait (STR), kinerja ujian pada materi terkait (LPR), dan kinerja ujian pada tujuan tertentu (PEG). Perangkat lunak RapidMiner digunakan untuk melakukan clustering pada data yang diperoleh dari UCI Repository. Hasilnya menunjukkan lima kelompok data yang menunjukkan pola belajar pengguna, dan indeks Davies-Bouldin digunakan untuk mengevaluasi kinerja clustering yang baik. Penelitian ini memberikan wawasan penting tentang hubungan antara waktu belajar, pengulangan materi, dan prestasi ujian. Selain itu, penelitian ini membantu dalam pembangunan sistem pembelajaran yang dapat disesuaikan dengan teknologi.

Kata Kunci : K-Means, Clustering, RapidMiner, Pemodelan Pengetahuan Pengguna, Pembelajaran Adaptif.

PENDAHULUAN

Dalam era digital saat ini, pemodelan pengguna telah menjadi aspek penting dalam berbagai aplikasi seperti pembelajaran adaptif, dan evaluasi kinerja. Pemodelan ini memungkinkan analisis data pengguna untuk memahami pola perilaku, kinerja, dan kebutuhan individu. Dataset yang digunakan dalam penelitian ini berkaitan dengan waktu belajar, pengulangan materi, serta peforma siswa pada ujian. Dataset ini memiliki lima *variable* utama: STG (waktu belajar untuk tujuan tertentu), SCG (jumlah pengulangan untu tujuan tertentu), STR (waktu belajar untuk materi terkait), LPR (kinerja ujian pada materi terkait), dan PEG (kinerja ujian pada tujuan tertentu).

Penlitian ini bertujuan utuk mengeksplorasi hubungan antara parameter yang tersedia dalam dataset, penelitian ini dapat membantu dalam pengembagan serta perancangan system pembelajaran yang lebih efektif dan adaptif. Selain itu, analisis ini juga memberikan wawasan tentang bagaimana faktor-faktor seperti waktu belajar dan kinerja ujian mempengaruhi hasil akhir pengguna. Dengan menggunakan Teknik stastistik dan *machine learning*, penelitian ini diharapkan dapat memberikan kontribusi yang signifkan dalam memahami pola pembelajaran pengguna dan medukung pengembangan solusi yang inovatif dalam dunia pendidika berbasis teknologi.

METODE Penelitian

1. **Data *Mining***

   Data mining merupakan aktivitas menemukan, mengumpulkan, dan menganalisis sejumlah besar informasi untuk mengidentifikasi pola, keterkaitan, atau data berharga yang bisa dipakai untuk pengambilan keputusan. Pengumpulan dan pengolahan data ini dapat dilakukan dengan bantuan perangkat lunak yang memanfaatkan analisis statistik, matematika, atau teknologi seperti rapid miner serta Kecerdasan Buatan (*AI*). Secara umum, ada beberapa metode dalam melakukan data mining, termasuk Asosiasi, Klasifikasi, Regresi, dan Klustering. Data mining memiliki tiga tujuan utama, yaitu sebagai alat untuk menjelaskan atau explanatory, untuk validasi atau confirmatory, dan untuk penjelajahan atau exploratory.

2. ***K-Means***

   K-Means merupakan metode pembelajaran mesin tanpa pengawasan yang digunakan untuk mengelompokkan data dan mengenali pola. Prosesnya dimulai dengan pemilihan sejumlah data awal (K) secara acak, kemudian data tersebut akan diubah-ubah sampai ditemukan pengelompokan yang paling tepat. Tujuan dari pengelompokan K-means adalah untuk membagi data ke dalam K kelompok (*cluster*) berdasarkan kesamaan tertentu, sehingga data dalam satu kelompok lebih mirip satu sama lain dibandingkan dengan data dalam kelompok lainnya. Berikut adalah ilustrasi rumus jarak Euclidean.

$$d(x, y) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}$$

Gambar 1 Euclidean Distance

   Keterangan:

   d : determinan (*Euclidean Distance)*

   x : titik pusat cluster

   y : data

   n : jumlah data

   i : data ke-

3. ***Cluster***

   Clustering adalah teknik analisis data yang digunakan untuk mengelompokan data atau objek berdasarkan kesamaan karakteristik atau fitru untuk menemukan pola atau struktur tersembunyi dalam dataset. Tujuan utama dari clustering adalah untuk membagi data ke dalam grup (*cluster*) sedemikian rupa hingga objek dalam satu grup lebih mirip satu sama lain dibandingkan grup lain. Clustering sering

digunakan dalam berbagai bidang seperti ilmu computer, statistika, dan biologi untuk menemukan pola dalam data.

4. **Rapidminer**

   Rapidminer adalah platfrom data science yang digunakan untuk menganalisa data secara keseluruhan. *Platfrom* ini menyediakan berbagai prosedur penambangan data dan pembelajaran mesin, termasuk ekstak, transformasi, dan beban data(ETL), pra-pemrosesan data, visualisasi, pemodelan prediktif, dan evaluasi. Rapid miner memberikan kemampuan analisis data yang mendalam dan luas, memungkinkan pengembangan model secara otomatis dari awal hingga akhir.

HASIL DAN PEMBAHASAN

1. **Pengumpulan Data**

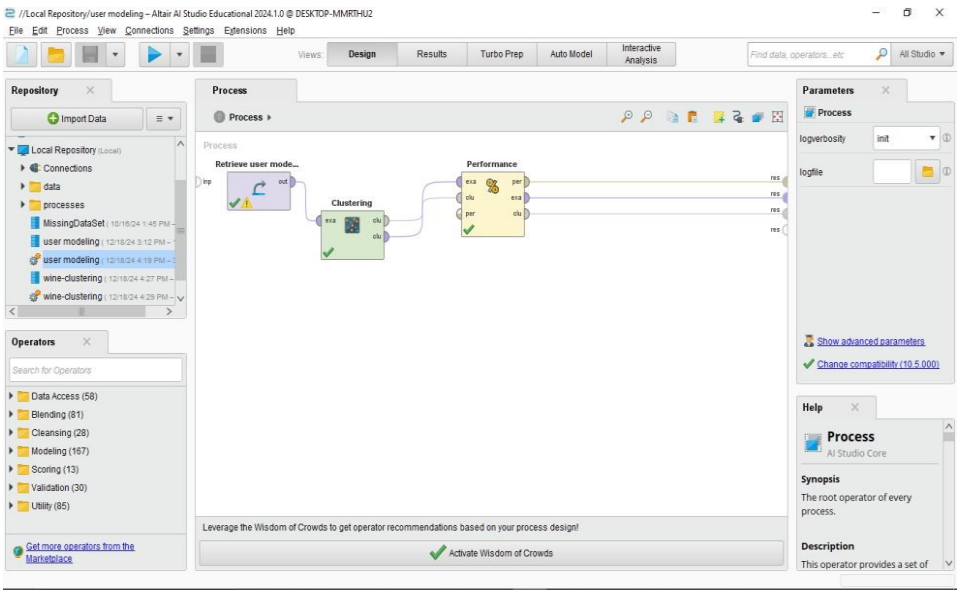
   Dataset yang digunakan dalam penelitian ini diperoleh dari *UCI Repository,* yaitu data *user know ladge* yang mencakup informasi pengguna terkait waktu belajar, pengulangan materi, dan pefroma ujian. Data ini memiliki lima *variable* utama *yaitu* STG, SCG, STR, LPR, dan PEG. Dataset ini diorganisasi dalam format terstruktru dan mencakup informasi yang relevan untuk analisis kinerja pembelajaran. Berikut adalah beberapa data yang akan ditampilkan:

Row No.	STG	SCG	STR	LPR	PEG
1	0	0	0	0	0
2	0.080	0.080	0.100	0.240	0.900
3	0.060	0.060	0.050	0.250	0.330
4	0.100	0.100	0.150	0.650	0.300
5	0.080	0.080	0.080	0.980	0.240
6	0.090	0.150	0.400	0.100	0.660
7	0.100	0.100	0.430	0.290	0.560
8	0.150	0.020	0.340	0.400	0.010
9	0.200	0.140	0.350	0.720	0.250
10	0	0	0.500	0.200	0.850
11	0.180	0.180	0.550	0.300	0.810
12	0.060	0.060	0.510	0.410	0.300
13	0.100	0.100	0.520	0.780	0.340
14	0.100	0.100	0.700	0.150	0.900
15	0.200	0.200	0.700	0.300	0.600

Gambar 2 Dataset user know ladge

2. *Procesing* Data

Rangkaian Tindakan yang dilakukan untuk mengonversi data yang belum diolah menjadi informasi yang bermanfaat. Tahapan ini mencakup beberapa Langkah utama, mulai dari pengumpulan data hingga analisis data. Pemodelan pada Rapidminer ditampilkan pada gambar berikut:



Gambar 3 Model Data Mining

3. Pemilahan atribut

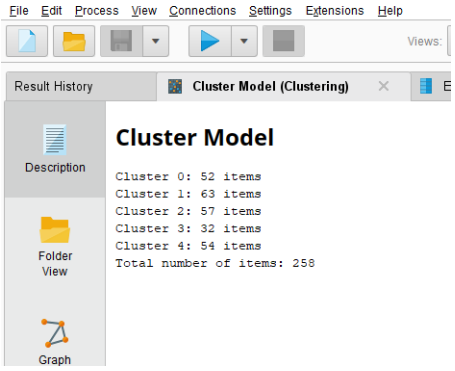
Dalam parameter untuk memilih atribut, jenis filter atribut dibagi lagi menjadi subset dan menentukan atribut mana yang akan digunakan.

a. Clustering

Penambahan port *Clustering k-means* dipilih dalam proses view untuk mengelompokakan kelas kelas pemodelan pengetahuan pengguna berdasarkan waktu belajar, pengulangan materi, dan pefroma ujian. Kelas yang akan dicari sebanyak lima cluster.

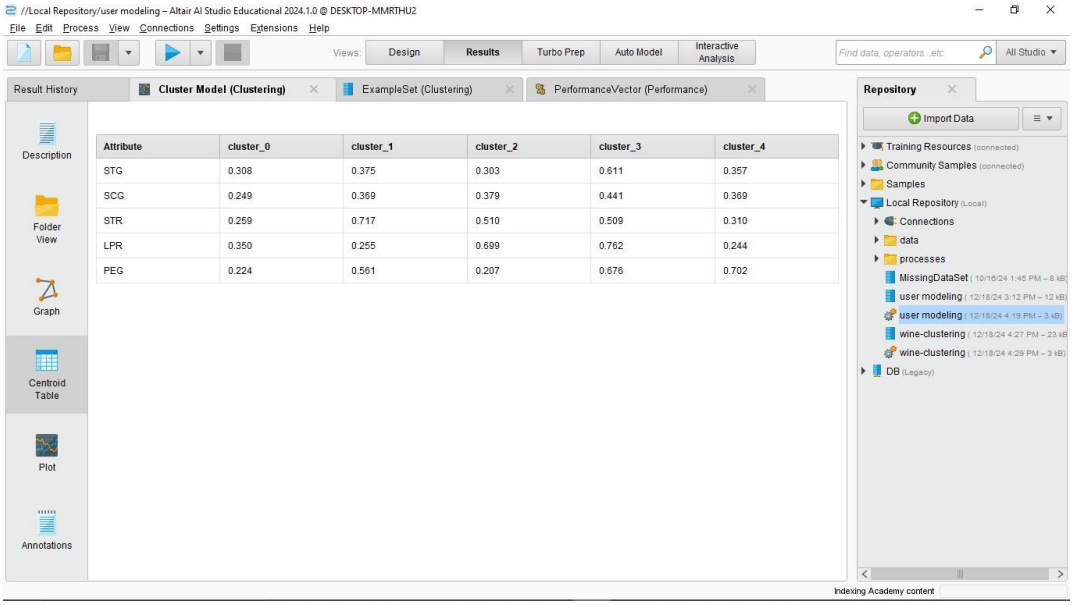
b. Perfomace

Perfomance yang digunakan adalah *cluster distance performance.* Lalu klik *run* untuk menampilkan hasil pengelolaan data.



Gambar 4 Cluster Model

4. *Centroid Table*

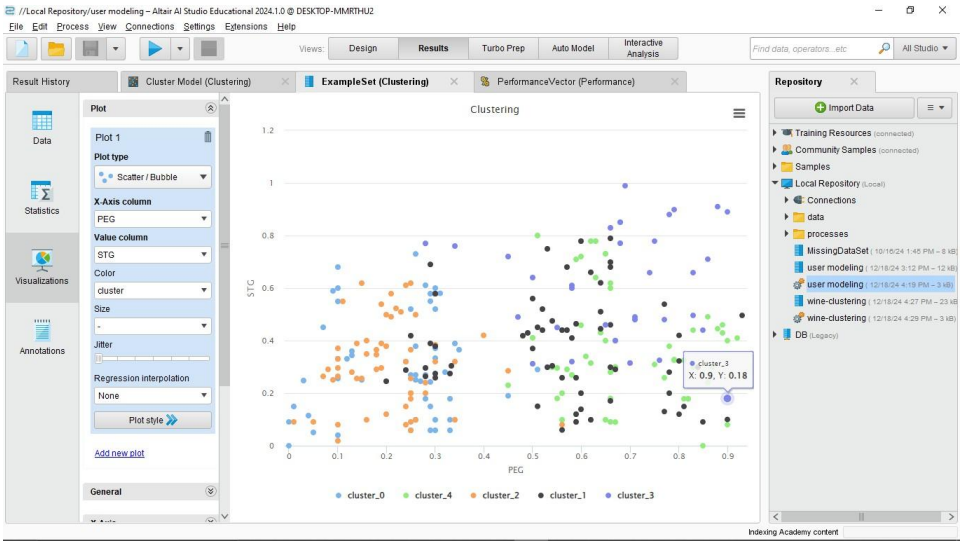


Attribute	cluster_0	cluster_1	cluster_2	cluster_3	cluster_4
STG	0.308	0.375	0.303	0.611	0.357
SCG	0.249	0.369	0.379	0.441	0.369
STR	0.259	0.717	0.510	0.509	0.310
LPR	0.350	0.255	0.699	0.762	0.244
PEG	0.224	0.561	0.207	0.676	0.702

Gambar 5 Centoid Table

Pada gambar ini, kita bisa mengetahui nilai rata-rata pada atribut kode STG, SCG, STR, LPR, PEG. Merupakan centroid table yang dihasilkan oleh rapidminer.

5. *Visualizations*

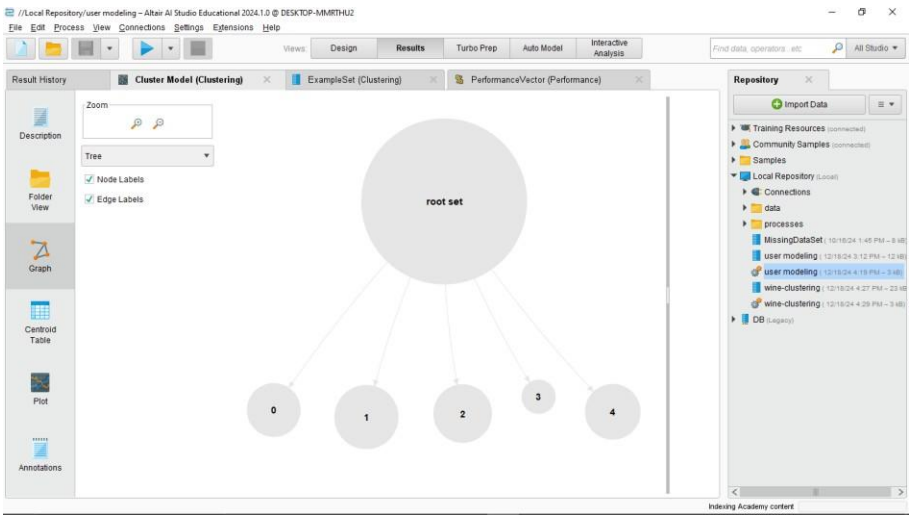


Gambar 6 visualizations

Gambar diatas, merupakan visualisasi data menggunakan *visualizations scalter/buble,* disana bisa melihat visualisasi data yang dihasilkan oleh rapidminer.

6. *Graph*

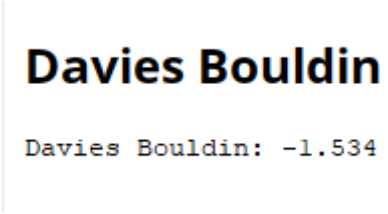
Pada gambar 7, model yang dihasilkan dari data yang dimiliki dan menciptakan *Rows* pengetahuan.



Gambar 7 Graph

7. ***Davies Bouldin***

Dapat disimpulkan bahwa proses klastering yang menggunakan algoritma k-means dalam studi ini cukup memuaskan, berdasarkan penilaian yang dilakukan dengan indeks *Davies-Bouldin* dan nilai yang terlihat pada gambar di bawah ini.



Gambar 8 Hasil Davies Bouldin

KESIMPULAN

Dalam penelitian ini, algoritma *K-Means* digunakan untuk memclusterkan data pengguna yang mencakup waktu belajar, pengulangan materi, dan prestasi ujian. Lima variabel utama (STG, SCG, STR, LPR, dan PEG) termasuk dalam dataset yang diambil dari *UCI Repository* untuk proses analisis. Perangkat lunak rapidminer digunakan untuk melakukan analisis ini. Dengan menggunakan algoritma *K-Means*, pengelompokan data dilakukan ke dalam lima klaster selama tahapan pemrosesan data. Hasil evaluasi clustering, berdasarkan nilai indeks *Davies-Bouldin*, menunjukkan kinerja yang baik. Visualisasi hasil menunjukkan distribusi data berdasarkan fitur tertentu, yang memberikan wawasan tentang pola belajar pengguna. Dengan memberikan pemahaman yang lebih baik tentang hubungan antara waktu belajar, pengulangan, dan prestasi ujian, penelitian ini berkontribusi pada pengembangan sistem pembelajaran adaptif. Ini dapat mendukung inovasi dalam pembelajaran berbasis teknologi.

DAFTAR PUSTAKA

Alvianatinova, Via, et al. "PENERAPAN ALGORITMA K-MEANS CLUSTERING DALAM PENGELOMPOKAN DATA PENJUALAN SUPERMARKET BERDASARKAN CABANG (BRANCH)." JATI (Jurnal Mahasiswa Teknik Informatika) 8.2 (2024): 1529-1535.

Kusuma, Prasetyo Arta, and Ada Udi Firmansyah. "Deteksi Penyebaran Penyakit Tuberkulosis dengan Algoritma K-Means Clustering Menggunakan Rapid Miner." Jurnal Teknologi Informatika dan Komputer 8.2 (2022): 41-54.

Lestari, Putri Dwi, and Mulyawan Mulyawan. "Datamining Pada Penjualan Air Bersih Di Spam Akidah Menggunakan Algoritma K-Means Clustering Menggunakan Rapidminer." JATI (Jurnal Mahasiswa Teknik Informatika) 7.1 (2023): 412-416.

Pamungkas, Tri Bayu, Siti Maesaroh, and Pebri Ardiansyah. "Implementasi Data Mining Pada Stok Penggunaan Barang Di Gmf Aeroasia Menggunakan Algoritma K-Means Clustering." Jurnal Ilmiah Sains dan Teknologi 7.2 (2023): 112-123.

Sari, Y. Ratna, et al. "Penerapan Algoritma K-Means Untuk Clustering Data Kemiskinan Provinsi Banten Menggunakan Rapidminer." CESS (Journal Comput. Eng. Syst. Sci., vol. 5, no. 2, p. 192, 2020, doi: 10.24114/cess. v5i2. 18519 (2020).