

Implementasi Algoritma Klasterisasi K-Medoids untuk Segmentasi Pengguna E-Ujian.Com

Anggyi Trisnawan Putra

Universitas Negeri Semarang, Indonesia
Email: anggy.trisnawan@mail.unnes.ac.id

Abstract

User behaviour management and analysis is very important in business. A good marketing strategy needs to be done so that the loyalty of old users is maintained. This can be initiated by segmenting users so that a good marketing strategy can be formulated. Customer segmentation can be done with the help of one of the data mining methods, which is clustering. In this study, k-medoids algorithm is used to cluster e-ujian.com users based on the behavioral data of each user. The first step will be analyzing data attributes that can be used. Next, the clustering process was carried out with the experimentally determined value of k. Finally, the cluster results will be evaluated using the Davies Bouldin Index (DBI) to determine the best number of clusters. The results showed that the value of $k = 4$ became the optimal number of clusters with a DBI value of 3,017.

Keywords: Clustering, k-medoids, Customer Segmentation

Abstrak

Pengelolaan dan analisis data perilaku pengguna menjadi hal yang sangat penting di dalam bisnis. Strategi pemasaran yang baik perlu dilakukan agar loyalitas pengguna lama tetap terjaga. Hal ini dapat diinisiasi dengan melakukan pengelompokan pengguna sehingga dapat diformulasikan strategi pemasaran yang baik. Pengelompokan pelanggan dapat dilakukan dengan bantuan salah satu metode data mining yaitu klasterisasi. Pada penelitian ini, algoritma k-medoids digunakan untuk melakukan klasterisasi pengguna e-ujian.com berdasarkan data-data perilaku dari masing-masing pengguna. Langkah pertama akan dilakukan analisis atribut data yang dapat digunakan. Selanjutnya, proses klasterisasi dilakukan dengan nilai k yang ditentukan secara eksperimental. Terakhir, hasil klaster akan dievaluasi menggunakan Davies Bouldin Index (DBI) untuk menentukan jumlah klaster yang terbaik. Hasil penelitian menunjukkan bahwa nilai $k = 4$ menjadi jumlah klaster optimal dengan nilai DBI sebesar 3,017.

Kata kunci: Klasterisasi, k-medoids, Segmentasi Pengguna

1. PENDAHULUAN

Aplikasi e-ujian.com merupakan software yang berkonsep SaaS (Software as a Service) B2B (Business to Business) yang membantu suatu instansi dalam menyelenggarakan tes secara online. Konsep B2B ini berarti layanan e-ujian.com memiliki pengguna dari kalangan instansi/entitas bisnis yang lainnya, bukan *customer* perorangan. Sebagai *startup* yang masih berkembang, pengelolaan dan analisis data perilaku pengguna menjadi hal yang sangat penting. Strategi pemasaran yang baik perlu dilakukan agar loyalitas pengguna lama tetap terjaga. Dalam penelitian yang dilakukan oleh [1], disebutkan bahwa *service quality* berpengaruh secara signifikan dan positif terhadap loyalitas pengguna layanan. Menurut [2], *service quality* memiliki beberapa dimensi, yaitu *reliability*, *responsiveness*, *assurance*, *empathy*, dan *tangibility*. Dalam salah satu dimensi tersebut, *responsiveness*, berarti seberapa kuat keinginan untuk membantu *customer* dan

menyediakan layanan secepat mungkin. Dalam kasus yang ada di e-ujian.com, sumberdaya manusia yang menjadi tulang punggung *customer service* belum sebanding dengan banyaknya *customer* yang ada. Sehingga, *service quality* menjadi hal yang rawan untuk tidak terpenuhi dengan baik.

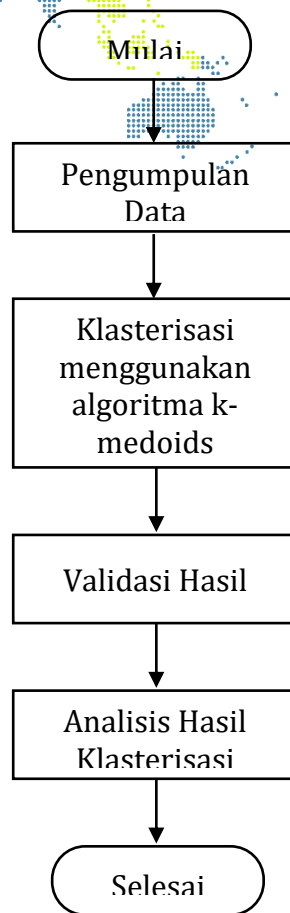
Banyaknya pengguna di e-ujian.com dapat disiasati dengan mengelompokkannya menjadi beberapa kelompok sesuai dengan perilaku atau kondisi datanya. Pengguna yang ada di satu kelompok akan memiliki kesamaan-kesamaan tertentu sehingga aksi yang dilakukan dapat digeneralisasi untuk kelompok tersebut. Tentunya hal ini berdampak pada penanganan *customer* yang lebih praktis, misalnya saja fokus pemberian reward ditujukan pada kelompok tertentu yang dinilai lebih potensial. Pengelompokan pengguna ini selanjutnya dapat lebih dianalisis kembali sehingga dapat diformulasikan strategi pemasaran yang lebih baik [3].

Pengelompokan pengguna ini dapat dilakukan dengan teknik data mining yang biasa disebut dengan klasterisasi (*Clustering*). Jain dalam [4] mengartikan *clustering* sebagai metode yang mengelompokkan data ke dalam banyak kelompok (klaster) berdasarkan kemiripan dan karakteristik data. Banyak algoritma klasterisasi yang dapat dipakai, seperti k-means, k-medoids, DBSCAN, dan lain sebagainya.

Penelitian terdahulu yang terkait dengan klasterisasi data *customer* diantaranya dilakukan oleh [5] yaitu segmentasi pelanggan CV Tita Jaya menggunakan algoritma k-means. Pada penelitian tersebut, data pelanggan berhasil disegmentasi menjadi enam klaster, dan hasil pengujian dengan perhitungan *silhouette coefficient* menghasilkan nilai 0,61. Penelitian serupa dilakukan Sutresno [6] menggunakan k-means untuk melakukan *clustering* data pelanggan, menghasilkan lima klaster. Dari lima klaster tersebut terdapat klaster 1, 3, dan 5 teridentifikasi sebagai *potential loyalist* sehingga pihak perusahaan dapat memberikan prioritas lebih ke pelanggan pada klaster tersebut. Sebagai *startup* B2B, e-ujian.com perlu melakukan segmentasi pengguna sehingga langkah kebijakan pengelolaan pengguna menjadi lebih efisien. Seperti pada [7], dilakukan segmentasi *retailer* untuk mengidentifikasi perilaku yang mirip dengan tujuan mengefisienkan strategi pemasaran. Setiap kelompok yang tersegmentasi dapat diatur strateginya sesuai dengan karakteristik kelompok tersebut. Pada penelitian ini digunakan algoritma k-medoids karena algoritma ini dikenal lebih dapat diandalkan, serta lebih cocok untuk data yang secara natural memang dapat dipisahkan dengan jelas [8].

2. METODOLOGI PENELITIAN

Metode yang digunakan pada penelitian ini dilakukan secara bertahap untuk menentukan segmentasi pengguna. Tahap-tahap yang ditempuh dapat digambarkan seperti Gambar 1.



Gambar 1. Tahapan penelitian

Penjelasan untuk masing-masing tahapan penelitian di atas akan dijelaskan sebagai berikut.

2.1. Pengumpulan Data

Tahap pengumpulan data ditempuh dengan cara mengakses basis data e-ujian.com yang berisi aktivitas pelaksanaan ujian dari seluruh lembaga yang terdaftar.

2.2. Klasterisasi Menggunakan Algoritma k-medoids

Proses klasterisasi menggunakan algoritma k-medoids dengan besaran jumlah klaster k yang akan ditentukan dalam beberapa kali percobaan. Semua proses klasterisasi dilakukan dengan bantuan aplikasi RapidMiner Studio. Pada aplikasi RapidMiner Studio sudah terdapat fitur segmentasi k-medoids yang tersedia dengan cukup memilihnya pada jendela operator.

2.3. Validasi Hasil Klastering

Validasi dilakukan dengan bantuan operator *Cluster Distance Performance* pada RapidMiner Studio, dengan memilih kriteria *Davies Bouldin* pada opsi perhitungannya. Beberapa nilai k akan dicoba dan

divalidasi hasilnya menggunakan perhitungan *Davies Bouldin Index* (DBI). Semakin kecil nilai DBI, maka semakin baik pula hasil klusterisasinya [4]. Sehingga nilai k yang terkait akan dipilih sebagai jumlah klaster optimal.

2.4. Analisis Hasil Klusterisasi

Pada tahap ini dilakukan analisis terhadap klaster optimal yang telah dihasilkan pada tahapan sebelumnya. Analisis ini dilakukan dengan mengamati hasil segmentasi pengguna sejumlah klaster optimal, sehingga karakteristik tiap klaster dapat ditentukan.

3. HASIL DAN PEMBAHASAN

3.1. Pengumpulan Data

Pengumpulan data dilakukan dengan mengakses basis data e-ujian.com yang terkumpul sejak tahun 2020 dan berisi tabel-tabel antara lain:

- 1) Tabel lembaga
- 2) Tabel riwayat pembelian paket poin
- 3) Tabel riwayat penyelenggaraan ujian
- 4) Tabel paket soal yang dibuat lembaga

Total data yang terkumpul dengan penggabungan beberapa tabel di atas adalah sebanyak 827 *records*. Penggabungan beberapa tabel ini menghasilkan atribut terpilih yang diharapkan akan menjadi patokan pengelompokan pengguna. Pada Tabel 1 disajikan atribut terpilih yang akan diproses ke tahap selanjutnya.

Tabel 1. Daftar atribut terpilih

No.	Nama Atribut	Kode	Keterangan	Tipe Data
1	id_lembaga	A	ID untuk keperluan identifikasi data lembaga	Numerik (ID)
2	jml_poin	B	Jumlah poin yang pernah dibeli oleh masing-masing lembaga	Numerik
3	jml_peserta	C	Jumlah peserta ujian yang ada di lembaga	Numerik
4	jml_sesi	D	Jumlah sesi pelaksanaan ujian yang pernah dibuat oleh lembaga	Numerik
5	total_pakai	E	Jumlah poin yang pernah dipakai untuk keperluan penyelenggaraan ujian	Numerik
6	max_sehari	F	Maksimum penggunaan poin dalam sehari yang pernah dilakukan oleh lembaga	Numerik
7	jml_ps	G	Jumlah paket soal yang dibuat oleh lembaga	Numerik
8	harga_poin	H	Harga poin rata-rata dari masing-masing lembaga	Numerik
9	max_pembelian	I	Maksimum pembelian poin yang dilakukan dalam sekali waktu	Numerik

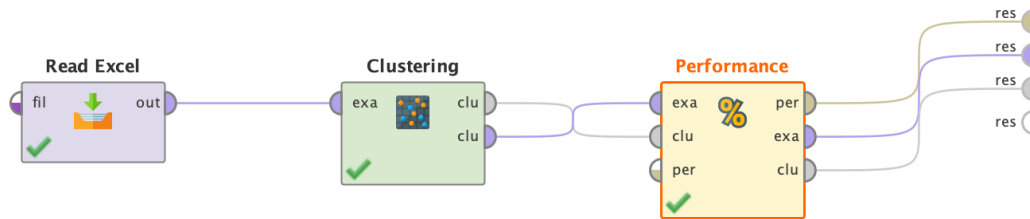
Sebagai gambaran, data hasil tahapan pengumpulan data sebagian disajikan pada Tabel 2.

Tabel 2. Sampel data

A	B	C	D	E	F	G	H	I
17	7455	386	911	5245	184	431	593	3000
33	650	70	96	279	14	44	1481	250
145	13250	618	1004	13444	451	383	517	2300
300	1000	958	12	862	740	7	1100	1000
364	250	28	64	50	9	39	701	250
381	1350	241	327	1206	59	187	1121	500
389	1219	226	194	1319	88	86	740	417
614	100	112	19	104	81	5	753	100
625	1410	114	59	747	73	44	624	1200
627	1500	668	45	711	119	18	748	1000
645	250	55	23	133	34	16	701	250
662	4500	332	910	3829	54	545	668	1000
675	49500	1986	2047	41401	812	855	530	16500
687	4150	4150	128	3936	1111	11	609	1600
742	13353	1994	667	10411	952	67	565	4990
759	56610	1547	3361	63017	1101	534	566	8000
760	1500	208	19	1440	177	13	600	750
786	2350	521	531	2095	170	155	804	1000
794	21500	15349	3168	15568	1433	51	652	6500
802	460	177	43	332	38	22	713	150
837	1091	161	54	600	48	17	627	741
863	100	10	1	3	3	1	753	100
881	250	117	91	221	31	34	1002	100
929	445	360	417	366	56	27	974	250
...	...	...	...	...	...	...	...	...
662	4500	332	910	3829	54	545	668	1000
9139	500	134	13	182	112	9	1170	500
9142	250	135	2	125	92	1	1050	250
9161	100	144	8	39	14	2	675	100
9166	500	278	7	259	220	7	975	500
9217	100	2	1	1	1	1	675	100
9245	1100	51	17	118	48	10	390	1000
9260	35	29	8	55	28	4	446	35
9266	500	488	2	247	247	1	975	500
9310	45	50	11	47	24	9	667	45
9367	100	81	6	79	79	2	1125	100

3.2. Klusterisasi Menggunakan Algoritma k-medoids

Proses klusterisasi dilakukan dengan bantuan RapidMiner Studio, dengan desain proses seperti pada Gambar 2.



Gambar 2. Desain proses klasterisasi

Proses dijalankan beberapa kali dengan nilai k yang berbeda-beda, mulai dari 2 s.d 6. Setiap prosesnya akan dihitung nilai DBI berdasarkan operator *Cluster Distance Performance*. Pada Tabel 3 di bawah ini disajikan hasil DBI untuk setiap percobaan nilai k .

Tabel 3. Hasil percobaan nilai k

No.	Nilai k	Hasil Klaster	Nilai DBI
1	2	Cluster 0: 29 items Cluster 1: 798 items	7,087
2	3	Cluster 0: 231 items Cluster 1: 372 items Cluster 2: 224 items	6,343
3	4	Cluster 0: 230 items Cluster 1: 231 items Cluster 2: 142 items Cluster 3: 224 items	3,017
4	5	Cluster 0: 230 items Cluster 1: 231 items Cluster 2: 118 items Cluster 3: 224 items Cluster 4: 24 items	3,892
5	6	Cluster 0: 77 items Cluster 1: 227 items Cluster 2: 83 items Cluster 3: 228 items Cluster 4: 147 items Cluster 5: 65 items	3,613

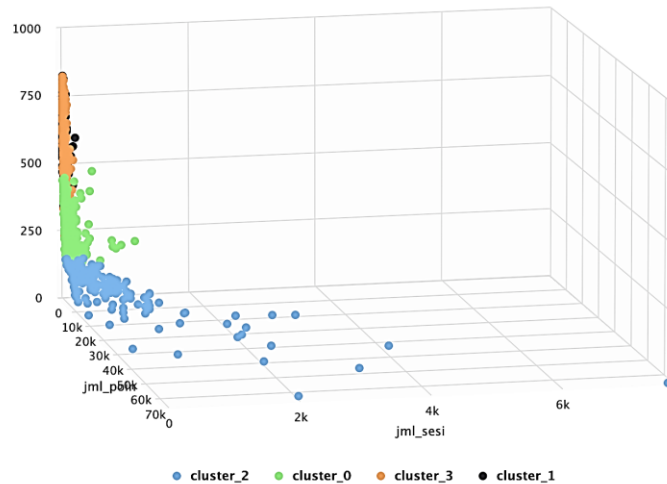
Dari hasil validasi menggunakan DBI yang tersaji pada Tabel 2, dapat disimpulkan bahwa nilai $k = 4$ merupakan nilai optimal. Segmentasi data pengguna e-ujian.com telah berhasil dilakukan dengan hasil klaster sebanyak 4. Pada Tabel 4 disajikan data *centroid* dari masing-masing klaster untuk nilai $k = 4$.

Tabel 4. *Centroid* dari masing-masing klaster

Atribut	cluster_0	cluster_1	cluster_2	cluster_3
total_pakai	657	2	3302	63
max_sehari	26	2	1634	12
max_pembelian	250	8	1700	10
jml_sesi	459	2	22	14
jml_ps	8	1	6	9
jml_poin	250	8	1760	10
jml_peserta	463	4	881	48

Atribut	cluster_0	cluster_1	cluster_2	cluster_3
harga_poin	1050	1153	1022	614

Sebagai gambaran, visualisasi data hasil klasterisasi dapat dilihat pada Gambar 3 di bawah ini.



Gambar 3. Visualisasi hasil klasterisasi dengan $k = 4$

3.3. Analisis Hasil Klasterisasi

Berdasarkan hasil klasterisasi optimal pada tahapan sebelumnya, dilakukan pengamatan terhadap hasil segmentasi, yaitu dengan melihat karakteristik dari anggota masing-masing klaster yang terbentuk. Hasil pengamatan ini menghasilkan label untuk masing-masing klaster sesuai dengan karakteristik datanya. Hasil pengamatan dapat dilihat pada Tabel 5.

Tabel 5. Hasil pengamatan data

Klaster	Jumlah Anggota	Label
0	230	<i>upper-regular-usage</i>
1	231	<i>low-usage</i>
2	142	<i>super-usage</i>
3	224	<i>lower-regular-usage</i>

Masing-masing label dibuat berdasarkan karakteristik data anggota klaster yang diamati, dan dapat dijelaskan pada Tabel 6.

Tabel 6. Penjelasan label tiap klaster

Label	Penjelasan
<i>low-usage</i>	Berisi pengguna yang tidak banyak menggunakan aplikasi, hanya lembaga yang memiliki aktivitas rendah (sekali pakai, hanya karena ada <i>event</i> tertentu) atau hanya coba-coba.
<i>lower-regular-usage</i>	Berisi lembaga yang menggunakan aplikasi secara berkala, tetapi dengan jumlah pemakaian yang tidak terlalu banyak. Lembaga yang ada di kategori ini banyak dari kalangan sekolah yang memiliki jumlah siswa berkisar antara 50 s.d 400.

Label	Penjelasan
<i>upper-regular-usage</i>	Berisi lembaga yang menggunakan aplikasi secara berkala dan dengan jumlah pemakaian yang cukup banyak. Lembaga yang ada di kategori ini didominasi oleh sekolah-sekolah yang memiliki jumlah siswa berkisar antara 500 s.d 1000 atau penyelenggara ujian yang secara rutin menyelenggarakan ujian.
<i>super-usage</i>	Berisi lembaga yang sering menggunakan aplikasi dengan jumlah pemakaian yang sangat besar, berisi sekolah-sekolah yang memiliki siswa berkisar di atas 1000, atau lembaga-lembaga besar (gabungan sekolah-sekolah atau tingkat kementerian) yang menyelenggarakan ujian dengan ribuan peserta sekali waktu.

Hasil analisis di atas menggambarkan dengan jelas bahwa klaster yang terbentuk dapat diinterpretasikan dengan baik, sesuai dengan karakteristik yang ada pada setiap klaster. Untuk setiap kelompok pengguna selanjutnya dapat diformulasikan strategi pemasaran yang sesuai dengan label tiap klaster, sehingga akan lebih efektif dan efisien.

4. SIMPULAN

Berdasarkan penelitian yang telah dilakukan, dapat disimpulkan bahwa klasterisasi telah berhasil dilakukan dalam rangka segmentasi data pengguna e-ujian.com. Klasterisasi dilakukan terhadap 827 baris data pengguna yang memiliki 8 atribut, menggunakan algoritma klustering k-medoids. Nilai k ditentukan secara *trial and error*, dan disimpulkan bahwa nilai $k = 4$ merupakan jumlah klaster yang optimal berdasarkan perhitungan DBI, yang menghasilkan nilai sebesar 3.017. Hasil klasterisasi optimal terdapat 4 klaster yang terbentuk, yaitu klaster 0 (*upper-lower-usage*) beranggotakan 230 lembaga, klaster 1 (*low-usage*) dengan anggota 131 lembaga, klaster 2 (*super-usage*) dengan anggota 142 lembaga, dan klaster 3 (*lower-regular-usage*) dengan anggota 224 lembaga. Terbentuknya 4 klaster ini diharapkan dapat mengefisienkan strategi bisnis e-ujian.com, dalam rangka memaksimalkan potensi marketing yang sesuai dengan karakteristik pengguna.

DAFTAR PUSTAKA

- [1] D. Jenderal *et al.*, "Pengaruh Service Quality Terhadap Loyalitas Dengan Kepuasan Pelanggan Sebagai Mediator Dalam Industri Farmasi B2b Di Sumatera Barat (Survey Pada Pelanggan Pt. Talang Gugun Sari Nusantara)," *Jurnal Ilmiah Mahasiswa Ekonomi Manajemen TERAKREDITASI PERINGKAT*, vol. 4, no. 3, pp. 589–601, 2020, [Online]. Available: <http://jim.unsyiah.ac.id/ekm>
- [2] M. A. A. Singaraj *et al.*, "Service Quality And Its Dimensions Chief Editor Editor Editorial Advisors Service Quality And Its Dimensions A Kowsalya 2," 2019. [Online]. Available: <https://www.researchgate.net/publication/333058377>

- [3] P. A. A. Pudjaningrum, C. S. Barkah, T. Herawaty, and L. Auliana, "Rumusan Program Membership, Poin Rewards dan Email Marketing untuk Meningkatkan Loyalitas Pelanggan: Studi pada Semanis Kamu Cafe," *Jurnal Administrasi Bisnis*, vol. 11, no. 1, pp. 21–30, Mar. 2022, doi: 10.14710/jab.v11i1.39815.
- [4] A. E. Ezugwu *et al.*, "A comprehensive survey of clustering algorithms: State-of-the-art machine learning applications, taxonomy, challenges, and future research prospects," *Engineering Applications of Artificial Intelligence*, vol. 110. Elsevier Ltd, Apr. 01, 2022. doi: 10.1016/j.engappai.2022.104743.
- [5] T. Ayu Rospricilia, S. Ayu Ithriah, and A. Anjani Arifiyanti, "segmentasi pelanggan menggunakan metode k-means clustering berdasarkan model rfm pada cv tita jaya," 2020.
- [6] S. Aprius, S. #1, A. I. #2, and E. Sedyono, "Metode K-Means Clustering dengan Atribut RFM untuk Mempertahankan Pelanggan," vol. 4, p. 433, 2018, doi: 10.28932/jutisi.v4i3.878.
- [7] P. Prasetya, M. Najib, F. Mahliza, K. Kunci, P. Dua-level, and P. K-means, "Retailer Segmentation as A Strategy for B2B Marketing using A Two-Level Clustering Analysis," Jun. 2022.
- [8] B. Lund and J. Ma, "A review of cluster analysis techniques and their uses in library and information science research: k-means and k-medoids clustering," *Performance Measurement and Metrics*, vol. 22, no. 3, pp. 161–173, Nov. 2021, doi: 10.1108/PMM-05-2021-0026.