

# Klasterisasi Penjawab Soal Berdasar Kualitas Jawaban pada Platform Brainly Menggunakan K-Means

Puji Winar Cahyo<sup>[1]\*</sup>, Landung Sudarmana<sup>[2]</sup>

Program Studi Informatika<sup>[1],[2]</sup>

Universitas Jenderal Achmad Yani Yogyakarta  
Sleman, Daerah Istimewa Yogyakarta

pwcahyo@gmail.com<sup>[1]</sup>, willerkasani@gmail.com<sup>[2]</sup>

**Abstract**— Brainly is a Community Question Answering (CQA) educational platform that makes it easy for users to find answers based on questions posed by students. Questions from students are often answered quickly by many answerers interested in the field being asked. The number of available answers is the choice of students to be able to receive answers and give a good rating to the answerer. Based on the number of good ratings, an answerer can be said to be an expert in certain subjects. Therefore, this research focuses on finding expert answering groups who have quality answers. K-means clustering is possible to group the answering data into two different clusters. The first cluster is expert users with ten respondents, and The second cluster is a non-expert cluster with 474 respondents. The expert cluster data is expected to help the questioner to be able to ask questions directly to the experts and obtain quality answers. Meanwhile, the number of clusters is determined based on the test results using a silhouette score that obtains a value of 0.971, with the optimal number of clusters being two clusters.

**Keywords**— clustering, brainly, k-means, question answer

**Abstrak**— Brainly merupakan platform edukasi Community Question Answering (CQA) yang memudahkan pengguna untuk mencari jawaban berdasarkan pertanyaan yang diajukan oleh siswa. Pertanyaan dari siswa seringkali dijawab dengan cepat oleh banyak penjawab yang memiliki kepedulian pada bidang yang sesuai. Banyaknya jawaban yang tersedia menjadi pilihan siswa untuk bisa menerima jawaban dan memberikan rating baik kepada penjawab. Melalui tingkat banyaknya rating baik yang dimiliki oleh penjawab pada suatu mata pelajaran tertentu, maka penjawab dapat dikatakan ahli pada mata pelajaran tersebut. Oleh karena itu penelitian ini berfokus pada pencarian kelompok penjawab ahli yang memiliki jawaban berkualitas. Menggunakan algoritma kluster k-means dapat mengelompokkan data penjawab kedalam 2 kluster yang berbeda yaitu kluster ahli dengan 10 penjawab dan kluster bukan ahli dengan 474 penjawab. Data kluster ahli tersebut diharapkan dapat membantu penanya untuk dapat menanyakan pertanyaan secara langsung pada ahlinya sehingga memperoleh jawaban yang berkualitas. Sementara itu penentuan banyaknya kluster didasarkan pada hasil pengujian menggunakan silhouette score memperoleh nilai 0.971 dengan jumlah kluster optimal adalah dua kluster.

**Kata Kunci**— kluster, brainly, k-means, tanya jawab

## I. PENDAHULUAN

Sulitnya siswa dalam menjawab pertanyaan soal yang diberikan sekolah memicu adanya keinginan siswa mencari jawaban melalui sumber yang mereka percaya. Sejalan dengan berkembangnya teknologi saat ini, sumber terpercaya terkait bidang pendidikan mulai dibangun. Sumber terpercaya dengan bentuk platform digital menjadi peluang untuk menciptakan cara menjawab soal dengan mudah dan cepat. Salah satu platform *education technology* yang berkembang pesat saat ini adalah Brainly. Brainly merupakan platform *question answer* yang didalamnya memberikan kebebasan pengguna untuk dapat menanyakan jawaban soal yang mereka anggap sulit untuk dikerjakan [1]. Pertanyaan yang diajukan tersebut nantinya akan dijawab oleh pengguna lain dilanjutkan dengan pemberian rating terhadap jawaban yang telah diberikan, sehingga proses jawab pertanyaan yang dilakukan dapat membentuk cara komunikasi interaktif dua arah. Cara komunikasi yang interaktif tersebut menarik banyaknya siswa dan penjawab untuk selalu diskusi dan menjawab pertanyaan melalui Brainly. Dari banyaknya pengguna aktif yang menggunakan Brainly maka terbentuk *Community Question Answering* (CQA) di bidang edukasi yang dapat membantu proses belajar siswa untuk bisa memahami jawaban pertanyaan secara lebih mendalam.

Secara umum CQA Merupakan tempat dimana para pengguna *platform* dapat melakukan *share* informasi dengan tujuan mencari jawaban terhadap suatu pertanyaan. Jawaban yang diperoleh akan dilakukan validasi keterkaitan dan kebenaran oleh penanya. Apabila jawaban yang diberikan oleh penjawab terindikasi benar, maka penjawab akan memperoleh reputasi baik. Banyaknya reputasi baik tersebut dapat menunjukkan bahwa sipenjawab merupakan penjawab yang kredibel [2]. Setiap *platform* CQA memiliki cara berbeda dalam memberikan penghargaan terhadap reputasi baik yang dihasilkan oleh seorang penjawab. *Platform StackOverflow* menggunakan cara reputasi pengguna dapat bertambah nilainya apabila penjawab memberikan jawaban yang sesuai dan jawaban dianggap diterima. Begitu sebaliknya reputasi baik akan berkurang apabila pertanyaan atau jawaban yang diajukan adalah pertanyaan atau jawaban yang sembarangan [3].

Platform Brainly mengakomodasi cara pemberian nilai yang berbeda dalam penentuan reputasi penjawab. Reputasi pengguna dapat dikatakan meningkat baik apabila penjawab dapat memberikan jawaban terhadap pertanyaan yang telah diajukan. Tidak hanya itu, nilai reputasi baik dapat bertambah apabila jawaban yang diberikan merupakan jawaban terbaik dari jawaban yang lain. Setiap nilai yang dihasilkan akan dikonversi menjadi poin yang dapat dikalkulasi jumlahnya. Hasil kalkulasi tersebut nantinya berpengaruh pada tingkat lencana yang diberikan oleh Brainly. Pemberian lencana pada pengguna Brainly memacu pengguna untuk berlomba mengumpulkan poin dengan cara menjawab pertanyaan sebanyak-banyaknya [4]. Keinginan pada banyaknya poin yang dikejar oleh penjawab berdampak pada banyaknya jawaban yang kurang berkualitas [5].

Secara mendasar jawaban yang berkualitas dapat diukur dari tingkat kepuasan penanya. Semakin tinggi tingkat kepuasan penanya maka jawaban dapat dikatakan berkualitas. Tidak hanya itu, dari banyaknya jawaban yang berkualitas maka penjawab dapat dikatakan sebagai pengguna ahli [6] yang memiliki profil baik [7]. Identifikasi terhadap penjawab ahli dengan profil baik tersebut dapat membantu penanya untuk dapat menanyakan permasalahan secara langsung kepada penjawab ahli. Didalam platform Brainly kelompok penjawab dengan *rating* tinggi dapat dilihat pada halaman awal di kolom pengguna tercerdas, akan tetapi kelompok pengguna tercerdas tersebut belum terbagi sesuai bidang keahlian masing-masing dalam menjawab. Pengelompokan pengguna ahli sesuai bidang keahliannya diperlukan untuk mempermudah penanya dalam memilih penjawab. Selain itu, hasil jawaban dari pertanyaan yang diajukan kepada pengguna ahli tersebut maka diharapkan dapat memperoleh jawaban yang berkualitas dan sesuai. Mengelompokan data pengguna ahli dapat dilakukan dengan menggunakan algoritma kluster berdasarkan pola parameter yang sama. Algoritma kluster k-means digunakan untuk pengelompokan data penjawab pada platform Brainly, karena dinilai memiliki tingkat akurasi yang tinggi dibanding algoritma kluster yang lain [8].

Pengelompokan pengguna ahli sesuai bidang keahliannya diperlukan untuk mempermudah penanya dalam menanyakan pertanyaan yang sesuai dengan bidang keahliannya. Dari pertanyaan yang diajukan kepada pengguna ahli tersebut maka diharapkan dapat memperoleh jawaban yang berkualitas dan sesuai.

## II. TINJAUAN PUSTAKA

### A. Penelitian pada *Community Question Answering*

Penelitian mengenai *Community Question Answering* (CQA) saat ini sudah sangat sering dilakukan terutama pada pemrosesan data teks. Data teks yang tersedia dan diolah umumnya berupa format pertanyaan dan jawaban. Jawaban dari pertanyaan yang berupa teks digunakan sebagai sumber data untuk membentuk model pemilihan jawaban terbaik. Melalui hasil pemilihan jawaban terbaik tersebut digunakan sebagai ukuran tingkat keahlian penjawab [9]. Selain itu penelitian lain dengan memanfaatkan data CQA digunakan untuk melatih model dalam pemilihan jawaban didalam sebuah paragraf dengan cara *semantic correlation* menggunakan *Recurrent*

*Convolutional Neural Network* (RCNN) dan similarity kata menggunakan *word embedding* yang saling terkait dapat menghasilkan tingkat akurasi yang tinggi yaitu 1.6% melebihi model pemilihan jawaban yang telah disubmit pada SemEval2015 [10].

Sementara itu, dalam kasus *Frequency Answer Question* (FAQ) biasanya penanya akan menanyakan jawaban yang dinilai sering ditanyakan. Pada kasus FAQ, biasanya permasalahan muncul apabila pertanyaan tersebut bukan salah satu pertanyaan yang berada pada kategori FAQ. Oleh karena itu untuk mengatasi permasalahan tersebut maka dapat menggunakan *deep learning* dengan tujuan melakukan analisis pada setiap pertanyaan yang akan diajukan. Menggunakan *Long Short Term Memory* (LSTM) dan *Conditional Random Fields* (CRF) terbukti dapat menaikkan tingkat akurasi pemilihan jawaban diluar FAQ [11]. Selain berfokus pada bentuk data yang berupa teks, pencarian jawaban terbaik dapat dilakukan melalui pengukuran tingkat penjawab terbaik. Pengukuran tingkat keahlian penjawab dapat dilihat dari banyaknya *rating* jawaban yang diterima oleh penanya, sehingga penjawab tersebut dapat dikatakan ahli pada bidangnya melalui jumlah *rating* tersebut. Pengelompokan penjawab dengan keahlian yang sejenis dapat dilakukan menggunakan metode kluster. Klusterisasi kelompok penjawab berdasarkan bidang keahlian menggunakan algoritma Fuzzy C-Means menghasilkan dua kluster kelompok pengguna yang berbeda, sesuai pola banyaknya *rating* yang dimiliki oleh setiap penjawab [12].

### B. Klusterisasi pada *Community Question Answering*

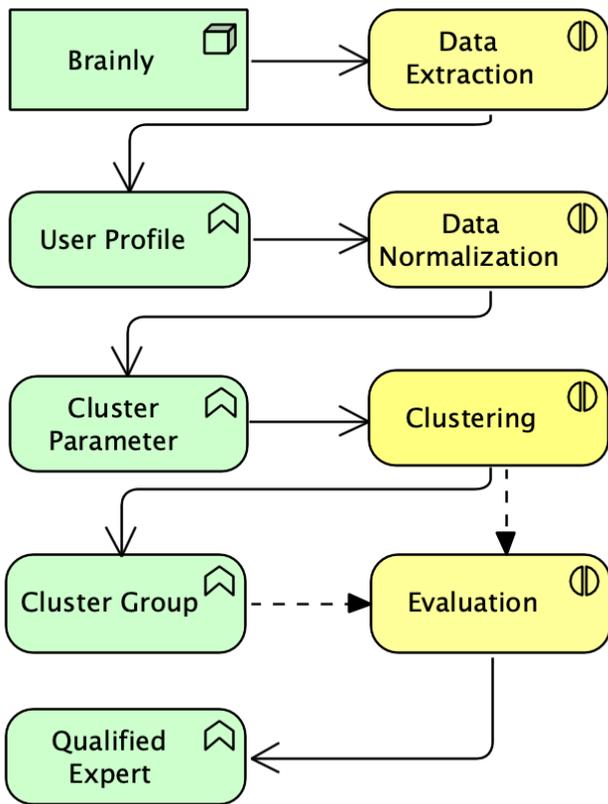
Metode klasifikasi seringkali dipakai pada pengelompokan data melalui proses pelatihan menggunakan data berlabel, data berlabel yang dimaksud adalah setiap data pasti akan memiliki identifikasi label yang menandakan data tersebut termasuk dalam kelompok tertentu. Berbeda dengan metode klusterisasi, metode klusterisasi adalah pengelompokan data tanpa melalui proses pelatihan dengan menggunakan data yang sudah diberi label [13] tujuan proses ini adalah mengelompokan data yang tersebar agar dapat dijadikan satu sesuai dengan kelompok tipe yang homogen [14]. Seperti penerapan algoritma k-means yang digunakan untuk pengelompokan data training pada *Visual Question Answering* (VQA) [15], data akan dilakukan eliminasi pada tingkat akurasi di bawah 70 persen untuk menghasilkan model yang baik.

Penelitian lain penerapan metode kluster untuk *Question Generation* pada *Question Answering* [16] yang bertujuan untuk melakukan segmentasi pertanyaan berdasar pada tingkat similaritas konteks yang ditanyakan, hasil dari segmentasi dapat mengelompokan pertanyaan dengan tingkat similaritas konteks yang sama. Disamping itu, implementasi segmentasi konteks juga dilakukan pada *Complex Factoid Question Answering* yang berfokus pada parafrase pertanyaan sejenis dengan memperhatikan keragaman leksikal dan sintaksis kata yang menyusun pertanyaan [17]. Hasil dari parafrase yang diterapkan dapat mengubah pertanyaan dengan susunan kata yang kurang jelas atau kurang lengkap menjadi pertanyaan yang mempunyai makna.

## III. METODE PENELITIAN

Penelitian ini menggunakan metode eksperimen yang

berfokus pada pengolahan data secara kluster. Pengolahan data secara kluster tersebut tentunya berdasar pada arsitektur yang telah dirancang sebelumnya. Hasil rancangan arsitektur berupa tahapan penelitian yang tergambar secara dan dapat dilihat pada Gambar 1. Tahapan penelitian dimulai dari data CQA yang berasal dari platform Brainly kemudian dilanjutkan pada proses *Data Extraction*, sehingga menghasilkan data *User Profile* yang kemudian dilakukan normalisasi data pada proses *Data Normalization*. Proses normalisasi data menghasilkan *Cluster Parameter* yang nantinya akan digunakan sebagai pola dasar pembentukan kluster pada proses *Clustering*. Proses *Clustering* akan membagi data kluster sesuai pola sejenis dengan parameter banyaknya *rating* dan jumlah jawaban. Sementara itu pada tahap evaluasi dilakukan proses pengujian algoritma kluster sesuai dengan banyaknya jumlah kluster yang telah dihasilkan pada Proses *Clustering*. Evaluasi algoritma yang dilakukan untuk mencari jumlah kluster optimal yang akan digunakan sebagai penentu penjawab yang memiliki kualifikasi penjawab ahli di bidangnya atau *Qualified Expert*.



Gambar 1. Arsitektur penelitian

A. *Data Extraction*

*Data extraction* merupakan proses dimana data profil pengguna brainly diperoleh. Proses ekstraksi data dimulai dengan menelusuri pengguna platform Brainly secara acak kemudian memulai pengambilan data secara otomatis menggunakan cara *web data extraction* [18]. Data yang diambil berupa detail profil yang melekat pada setiap pengguna, diantaranya adalah *username*, poin, lencana dan jumlah jawaban tercerdas mata pelajaran tertentu, secara antarmuka

dapat dilihat pada Gambar. 2 dan Gambar. 3.

**Sutr1sn0**  
 364.929 poin  
 Brainly Doctor    Jenius  
 Jawaban: 11.9 rb    Tercerdas: 1.8 rb    Terima Kasih: 57.9 rb

Gambar. 1. Profile pengguna Brainly

Tercerdas dalam

- Matematika**  
10.689 Jawaban (1.584 Tercerdas)
- Fisika**  
440 Jawaban (60 Tercerdas)
- Biologi**  
130 Jawaban (11 Tercerdas)
- B. Inggris**  
102 Jawaban (13 Tercerdas)

Gambar. 2. Jawaban tercerdas

Pada Gambar. 1 dapat dilihat bahwa **Sutr1sn0** memiliki lencana Jenius dikarenakan **Sutr1sn0** memiliki lebih dari 15 ribu poin dan lebih dari 50 jumlah jawaban tercerdas. Tidak hanya itu, setiap pengguna *platform* Brainly memiliki lencana berbeda-beda tergantung tingkat poin dan jumlah jawaban tercerdas yang dimiliki. Pada Gambar 2 dapat dilihat bahwa setiap pengguna platform Brainly masing-masing memiliki total jawaban tercerdas dari setiap mata pelajaran. Dari data yang dapat dilihat dari kedua antar muka tersebut maka data profil Sutr1sn0 dapat diambil dan dimasukkan kedalam database mongoddb sehingga membentuk format seperti Tabel 1.

TABEL 1. DATA PROFILE

<i>username</i>	<i>lencana</i>	<i>Jumlah Jawaban (Bahasa Inggris)</i>	<i>Jumlah Rating Tercerdas (Bahasa Inggris)</i>
Sutr1sn0	Jenius	102	13
Arkan5283	Jenius	26	7
Cadillac	Si Hebat	23	4

**B. Data Normalization**

Data Normalization merupakan tahapan dimana data yang sudah tersedia diproses menjadi data yang terstandarisasi dengan format tertentu untuk dapat dilanjutkan pada tahap selanjutnya [19]. Dari data profil yang telah disimpan dalam database maka dapat dilakukan normalisasi data untuk dijadikan parameter yang nantinya akan diteruskan pada tahap clustering. Normalisasi data yang dilakukan adalah mencari jumlah jawaban yang tidak memiliki rating tercerdas pada setiap mata pelajaran dengan rumus:

$$Answer\ No\ Rating = Total\ Answer - Smartest\ Answer \quad (1)$$

Dapat dilihat pada Persamaan (1) Answer No Rating adalah total jawaban yang tidak mendapatkan rating, Total Answer merupakan total jawaban, Smartest Answer adalah total jawaban dengan rating tercerdas. Dari hasil proses normalisasi data maka dapat menghasilkan data seperti pada Tabel 2.

TABEL 2. DATA NORMALISASI

<i>username</i>	<i>lencana</i>	<i>Jumlah Rating Tercerdas (Bahasa Inggris)</i>	<i>Jumlah Tidak dengan Rating Tercerdas (Bahasa Inggris)</i>
Sutr1sn0	Jenius	13	89
Arkan5283	Jenius	7	19
Cadillac	Si Hebat	4	19

Pada Tabel 2. Dapat dilihat bahwa Sutrlsno memiliki jumlah rating tercerdas bahasa inggris sebanyak 13 jawaban dan memiliki jumlah jawaban tidak dengan rating sebanyak 89. Nilai jumlah tidak dengan rating tersebut didasarkan dari Persamaan (1) yaitu melalui pengurangan jumlah jawaban 102 - 13 = 89. Sementara itu penjawab dengan nama pengguna Arkan5283 memiliki jumlah rating tercerdas 13 dan jumlah tidak tercerdas 19, begitu juga dengan penjawab yang lainnya mendapatkan nilai normalisasi dengan cara yang sama.

**C. Clustering**

Tahapan Clustering merupakan proses dimana hasil data normalisasi akan dilakukan pengelompokan sesuai dengan tingkat homogenitas data yang tersebar. Algoritma cluster yang diterapkan pada penelitian ini menggunakan k-means clustering yaitu dengan melakukan optimasi sebuah fungsi untuk menghitung distance space antara objek dengan centroid (titik tengah) kluster [20]. Pengukuran distance space pada penelitian ini menggunakan algoritma euclidean distance [21]. Rumus algoritma k-means dapat dilihat melalui Persamaan (2) [22].

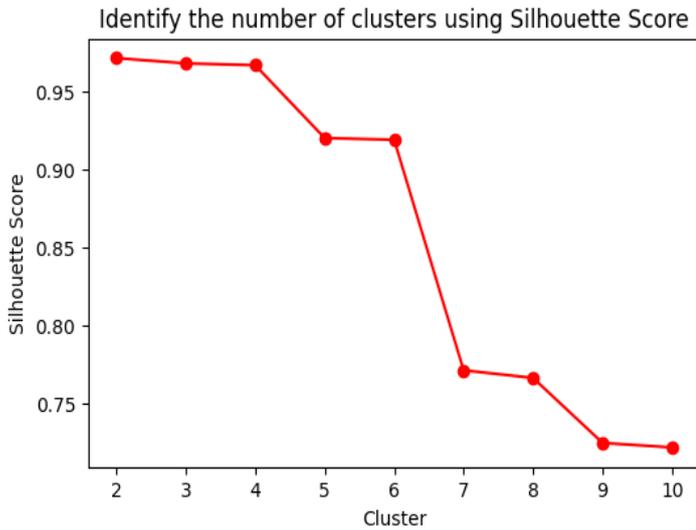
$$d(x_j, c_j) = \sqrt{\sum_{j=1}^n (x_j - c_j)^2} \quad (2)$$

Dilihat dari Persamaan (2) bahwa d merupakan jarak, n adalah banyaknya objek, j merupakan urutan data ke 1 sampai ke n, x<sub>j</sub> merupakan parameter ke j terhadap x dan c<sub>j</sub> merupakan centroid parameter ke j. Secara umum langkah algoritma k-means adalah

- 1) Menentukan banyaknya kluster (k);
- 2) Menentukan nilai titik centroid;
- 3) Analisis apakah centroid berubah berdasarkan mean setiap parameter objek (data) ? apabila ya, maka ubah centroid menjadi centroid baru, apabila tidak maka selesai;
- 4) Hitung distance space (jarak kedekatan) objek dengan centroid;
- 5) Kelompokkan objek berdasarkan kedekatan objek dengan centroid.

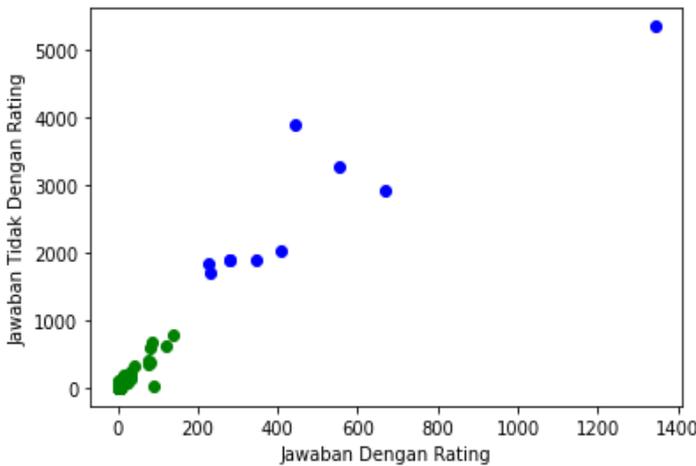
**IV. HASIL DAN PEMBAHASAN**

Pada penelitian yang telah dilakukan dengan menggunakan algoritma k-means klustering dapat menghasilkan jumlah kluster penjawab dengan jawaban yang berkualitas. Kelompok penjawab tersebut didasarkan dari banyaknya jumlah jawaban yang diberi rating tercerdas dan jumlah jawaban yang tidak mendapatkan rating. Kelompok penjawab tersebut dikelompokkan sesuai dengan jarak titik tengah masing-masing kluster. Titik tengah kluster yang didapat merujuk pada jumlah kluster optimal yang didapat dari hasil pengujian. Penentuan jumlah kluster optimal pada penelitian ini menggunakan silhouette score. Pengujian dilakukan secara acak dengan pengulangan sebanyak 10 kali jumlah kluster, hasil pengujian menghasilkan grafik yang dapat dilihat sesuai Gambar 3.



Gambar. 3. Silhouette score pada mapel bahasa inggris

Grafik pengujian dapat dilihat pada Gambar. 3. Pengujian dilakukan sebanyak 9 kali sehingga mendapatkan nilai jumlah kluster optimal yaitu 2 kluster dengan nilai 0.971, sedangkan untuk 3 kluster berada pada nilai 0.966, untuk 4 kluster mendapatkan nilai 0.967 kemudian menurun sampai jumlah kluster sebanyak 10. Dari banyaknya jumlah kluster optimal yang telah ditentukan, maka dapat dilakukan visualisasi kedalam 2 kluster yang memiliki tingkat homogenitas data yang sesuai, antarmuka tersebut dapat dilihat pada Gambar 4.



Gambar. 4. Sebaran kluster penjawab mapel bahasa inggris

Dapat dilihat dari Gambar. 4 kluster penjawab dengan warna hijau merupakan kelompok pengguna dengan kualitas jawaban yang tidak terlalu diperhitungkan. Sedangkan Kluster penjawab dengan warna biru merupakan kelompok pengguna dengan jawaban yang dinilai berkualitas. Kualitas tersebut didasarkan dari perbandingan jumlah jawaban *rating* tercerdas dengan jumlah jawaban yang tidak mempunyai *rating*. Segmentasi data penjawab dengan menggunakan k-means dapat membagi 2 grup kluster yang berbeda berdasarkan pola parameter yang sejenis, oleh karena itu hasil segmentasi data

dalam dua kluster dapat dilihat pada Tabel 3.

TABEL 3. DATA KLAS TER

<i>username</i>	<i>Jumlah Rating Tercerdas (Bahasa Inggris)</i>	<i>Jumlah Tidak dengan Rating Tercerdas (Bahasa Inggris)</i>	<i>Klaster</i>
Merlia42	1344	5348	1
darkzcliff	668	2927	1
Futaba21	556	3266	1
.....			
vy19	139	776	2
syifajs	122	626	2
RenAlEzhaben	86	674	2

Dapat dilihat dari Tabel 3 segmentasi kluster pertama dan kluster kedua terletak pada selisih data masing-masing parameter kluster. Titik centroid kluster pertama adalah [5, 31] apabila diukur tingkat kedekatannya maka akan menghasilkan 10 data penjawab yang memiliki tingkat homogenitas yang sama. Sedangkan titik centroid pada kluster kedua adalah [479, 2669] dengan total jumlah 474 data penjawab.

## V. KESIMPULAN

Berdasarkan penelitian yang dilakukan maka kelompok penjawab dengan tingkat jawaban berkualitas ditentukan dengan tingginya jumlahnya jawaban yang diberi *rating* tercerdas oleh penjawab. Selain itu paramater pendukung lain adalah banyaknya penjawab dalam menjawab pertanyaan pada suatu mata pelajaran tertentu, sementara pada kasus ini adalah bahasa inggris. Data kelompok penjawab pada kluster pertama dapat digunakan sebagai kelompok penjawab yang memiliki tingkatan ahli pada bidang mata pelajaran bahasa inggris. Sedangkan data kelompok pada kluster kedua merupakan data penjawab dengan kredibilitas menjawab soal belum diperhitungkan. Kondisi belum diperhitungkan tersebut karena dari jumlah jawaban dan *rating* tercerdas kelompok penjawab pada kluster kedua memiliki selisih jauh dengan kluster pertama.

Pada segi pemrosesan, penentuan jumlah banyaknya kluster sudah melalui proses perhitungan *silhouette score* terlebih dahulu, dengan menghasilkan nilai tertinggi yaitu 0.971 pada jumlah kluster optimal sebanyak 2 kluster, jumlah kluster optimal tersebut dapat diterapkan algoritma k-means *clustering*. Penerapan jumlah optimal kluster pada algoritma k-means klastering menghasilkan masing-masing titik tengah (*centroid*) kluster pertama adalah [5, 31] dan *centroid* kluster kedua adalah [479, 2669]. Data yang terbagi kedalam dua kluster tersebut digunakan sebagai penentu kualitas penjawab didalam platform Brainly.

UCAPAN TERIMAKASIH

Sumber dana penelitian ini berasal dari Hibah Internal Program Studi Informatika, Fakultas Teknik dan Teknologi Informasi Universitas Jenderal Achmad Yani Yogyakarta. Data yang digunakan atas kerjasama Fakultas Teknik dan Teknologi Informasi Universitas Jenderal Achmad Yani Yogyakarta dengan Brainly, untuk itu tim peneliti mengucapkan banyak terimakasih atas dukungan yang diberikan sehingga penelitian ini dapat terselesaikan.

DAFTAR PUSTAKA

- [1] Bramastia and E. K. Purnama, "Pengaruh Penggunaan Brainly Terhadap Hasil Belajar Siswa," *J. EPISTEMA*, vol. 2, no. 1, 2021.
- [2] M. Neshati, Z. Fallahnejad, and H. Beigy, "On dynamicity of expert finding in community question answering," *Inf. Process. Manag.*, vol. 53, no. 5, pp. 1026–1042, 2017.
- [3] A. Diyanati, B. S. Sheykhahmadloo, S. M. Fakhrahmad, M. H. Sadredini, and M. H. Diyanati, "A proposed approach to determining expertise level of StackOverflow programmers based on mining of user comments," *J. Comput. Lang.*, vol. 61, no. August, p. 101000, 2020.
- [4] R. Gharibi and M. Malekzadeh, "Gamified Incentives: A Badge Recommendation Model to Improve User Engagement in Social Networking Websites," *Int. J. Adv. Comput. Sci. Appl.*, vol. 8, no. 5, pp. 272–278, 2017.
- [5] S. Yanovsky, N. Hoernle, O. Lev, and K. Gal, "One size does not fit all: A study of badge behavior in stack overflow," *J. Assoc. Inf. Sci. Technol.*, vol. 72, no. 3, pp. 331–345, 2021.
- [6] L. T. Le, C. Shah, and E. Choi, "Assessing the quality of answers autonomously in community question–answering," *Int. J. Digit. Libr.*, vol. 20, no. 4, pp. 351–367, 2019.
- [7] P. W. Cahyo and M. Habibi, "Entity Profiling to Identify Actor Involvement in Topics of Social Media Content," *IJCCS (Indonesian J. Comput. Cybern. Syst.)*, vol. 14, no. 4, p. 417, 2020.
- [8] P. W. Cahyo and L. Sudarmana, "A Comparison of K-Means and Agglomerative Clustering for Users Segmentation based on Question Answerer Reputation in Brainly Platform," *Elinvo (Electronics, Informatics, Vocat. Educ.)*, vol. 6, no. 2, pp. 166–173, 2021.
- [9] S. Lyu, Y. Wang, W. Ouyang, H. Shen, and X. Cheng, "What we vote for? Answer selection from user expertise view in community question answering," *Web Conf. 2019 - Proc. World Wide Web Conf. WWW 2019*, pp. 1198–1209, 2019.
- [10] X. Zhou, B. Hu, Q. Chen, and X. Wang, "Recurrent convolutional neural network for answer selection in community question answering," *Neurocomputing*, vol. 274, no. 2017, pp. 8–18, 2018.
- [11] Y. Xiang, Q. Chen, X. Wang, and Y. Qin, "Answer Selection in Community Question Answering via Attentive Neural Networks," *IEEE Signal Process. Lett.*, vol. 24, no. 4, pp. 505–509, 2017.
- [12] P. W. Cahyo, K. Kusumaningtyas, and U. S. Aesy, "A User Recommendation Model for Answering Questions on Brainly Platform," *J. Infotel*, vol. 13, no. 1, pp. 7–12, 2021.
- [13] N. L. Sowah, Q. Wu, and F. Meng, "A classification and clustering method for tracking multiple objects," *2018 IEEE 8th Annu. Comput. Commun. Work. Conf. CCWC 2018*, vol. 2018-Janua, pp. 537–544, 2018.
- [14] P. W. Cahyo and M. Habibi, "Clustering followers of influencers accounts based on likes and comments on Instagram Platform," *IJCCS (Indonesian J. Comput. Cybern. Syst.)*, vol. 14, no. 2, pp. 199–208, 2020.
- [15] K. Kafle, M. Yousefhusien, and C. Kanan, "Data augmentation for visual question answering," *INLG 2017 - 10th Int. Nat. Lang. Gener. Conf. Proc. Conf.*, pp. 198–202, 2017.
- [16] N. Duan, D. Tang, P. Chen, and M. Zhou, "Question generation for question answering," *EMNLP 2017 - Conf. Empir. Methods Nat. Lang. Process. Proc.*, pp. 866–874, 2017.
- [17] A. Abujabal, R. S. Roy, M. Yahya, and G. Weikum, "COMQA: A community-sourced dataset for complex factoid question answering with paraphrase clusters," *NAACL HLT 2019 - 2019 Conf. North Am. Chapter Assoc. Comput. Linguist. Hum. Lang. Technol. - Proc. Conf.*, vol. 1, pp. 307–317, 2019.
- [18] P. W. Cahyo, M. Habibi, A. Priadana, and A. B. Saputra, "Analysis of Popular Hashtags on Instagram Account The Ministry of Health," vol. 34, no. Ahms 2020, pp. 270–273, 2021.
- [19] E. Ogasawara, L. C. Martinez, D. De Oliveira, G. Zimbrão, G. L. Pappa, and M. Mattoso, "Adaptive Normalization: A novel data normalization approach for non-stationary time series," *Proc. Int. Jt. Conf. Neural Networks*, 2010.
- [20] M. D. J. Bora and D. A. K. Gupta, "Effect of Different Distance Measures on the Performance of K-Means Algorithm: An Experimental Study in Matlab," vol. 5, no. 2, pp. 2501–2506, 2014.
- [21] R. I. Fajriah, H. Sutisna, and B. K. Simpony, "Perbandingan Distance Space Manhattan Dengan Euclidean Pada K-Means Clustering Dalam Menentukan Promosi," *Inform. Bina Sarana Bsi, Univ.*, vol. 4, no. 1, pp. 36–49, 2019.
- [22] P. W. Cahyo, "Klasterisasi Tipe Pembelajaran Sebagai Parameter Evaluasi Kualitas Pendidikan Di Perguruan Tinggi," *Teknomatika*, vol. 11, no. 1, pp. 49–55, 2018.