PROSIDING Seminar Nasional SISFOTEK ISSN 2597-3584 (Media Online)
Sistem Informasi dan Teknologi
website : www.seminar.iaii.or.id | email : seminar@iaii.or.id

Sistem Segmentasi Program Talk Show Berdasarkan Media Sosial Twitter Menggunakan Metode K-Medoids Clustering

Kharisma Jevi Shafira Sepyanto[1], Yulison Herry Chrisnanto [2], Fajri Rakhmat Umbara[3]
[123]Informatika, Fakultas Sains dan Informatika, Universitas Jenderal Achmad Yani
Jl. Terusan Jenderal Sudirman, Cimahi, Jawa Barat 40285
kharismashafiraa@gmail.com[1], y.chrisnanto@gmail.com[2], fajri.umbara@gmail.com[3]

Abstract

Innovations on a talk show on television can be a threat. Audience will be divided into groups so that it can make a downgrade rating program. Program ratings affect companies that will use advertising services. Television companies will go bankrupt. The biggest source of income is sales of advertising services. One way to overcome them can be analyzed in public opinion. The results of the analysis can provide information about the attractiveness of the community towards the program. But the analysis process takes a long time and can be done only by a competent person so another process is needed to get the results of the analysis that is fast and can be done by anyone. In this study using K-Medoids Clustering in the process of identifying public opinion. The clustering process known as unsupervised learning will be combined with the labeling process. The previous episode's tweet data will be labeled and then used to obtain the predicted labels from other cluster members. Labels consist of three types, namely 1) theme, 2) resource persons, and 3) programs. Before going through the clustering stage, the tweet data will go through the text preprocessing stage then transformed into a numeric form based on the appearance of the word. Transformation data will be clustered by calculating proximity using Cosine Similarity. Labels from the Medoids cluster will be used on unlabeled tweet data. The cluster results were tested using the Silhouette Coefficient method to get 0.19 results. However, this method successfully predicted public opinion and achieved an accuracy of 80%.

Keywords: Twitter social media segmentation, k-medoids clustering, cosine similarity, data transformation, silhouette coefficient, rating.

Abstrak

Inovasi yang hadir dalam acara *talk show* dapat menjadi ancaman dalam persaingan pertelevisian Indonesia. Menurunnya jumlah penonton program televisi menyebabkan penurunan *rating* pada program maupun stasiun televisi. Penurunan *rating* ini berpengaruh pada perusahaan yang berminat menggunakan jasa periklanannya. Apabila tidak segera diatasi perusahaan televisi swasta akan bangkrut karena sumber pendapatan terbesar adalah penjualan dari jasa periklanan. Salah satu cara mengatasinya dapat dilakukan analisis pada opini masyarakat. Hasil analisis dapat memberikan informasi gambaran masyarakat terhadap program. Namun proses analisis membutuhkan waktu yang panjang dan hanya bisa dilakukan oleh orang yang berkompeten sehingga diperlukan proses lain agar analisis opini dapat menghasilkan analisis yang cepat oleh siapa pun. Pada penelitian ini menggunakan K-Medoids *Cluster*ing dalam proses identifikasi opini masyarakat. Proses klasterisasi yang dikenal sebagai *unsupervised learning* akan dikombinasikan dengan proses pelabelan. Data tweet episode sebelumnya akan diberi label kemudian digunakan untuk memperoleh label yang diprediksi dari anggota *cluster* lainnya. Label terdiri dari tiga jenis yaitu 1)tema, 2)narasumber dan 3)program. Sebelum melewati tahap klasterisasi, data tweet akan melalui tahap *preprocessing* teks kemudian di transformasi menjadi bentuk numerik berdasarkan kemunculan kata. Data transformasi akan di klasterisasi dengan perhitungan kedekatan menggunakan Cosine Similarity. Label dari medoids *cluster* akan digunakan pada data tweet yang tidak berlabel. Hasil klaster di uji dengan metode Silhouette Coefficient mendapatkan hasil 0.19. Namun, metode ini berhasil memprediksi opini masyarakat dan mencapai akurasi sebesar 80%.

Kata kunci: segmentasi media sosial twitter, k-medoids clustering, cosine similarity, transformasi data, silhouette coefficient, rating.

1. Pendahuluan

Data yang selalu bertambah setiap detiknya dapat dimanfaatkan untuk memperoleh informasi. Proses olah data atau data mining dapat dilakukan dengan berbagai cara, salah satunya klasterisasi. Klasterisasi adalah metode pengelompokan data berdasarkan tingkat kemiripan karakteristik antar satu sama lainnya. Pada metode klasterisasi *partitional clustering* terdapat algoritma K-Means *Cluster*ing, K-Medoids *Cluster*ing (PAM) dan CLARA. Pada bidang efisiensi energi dalam mencari dan menganalisis periodisitas energi menggunakan algoritma K-Medoids *Cluster*ing dan CLARA, algoritma K-Medoids *Cluster*ing menjadi algoritma terbaik dalam kasus perhitungan matriks

menggunakan Euclidean Distance [1]. Pada penerapan 10.000 transaksi KEEL dengan algoritma K-Means dan K-Medoids, hasil menunjukkan K-Medoids lebih unggul dalam waktu eksekusi dan tidak senstitif terhadap *noise* [2]. Banyak bidang lain yang menggunakan klasterisasi dalam mengolah data.

Klasterisasi dapat diterapkan pada bidang *broadcasting*. Analisis terhadap opini masyarakat berhasil menghasilkan informasi yang diperlukan. Sebuah penelitian mengolah data hingga menghasilkan informasi gaya hidup masyarakat [3]. Klasterisasi dengan algoritma K-Means mendapatkan informasi keputusan produksi program hingga penentuan jadwal tayang sehingga dapat mempertahankan *rating* dari program [4]. Hal ini membuat para pengiklan ikut tertarik dalam menggunakan jasa periklanan di TV. Mereka beranggapan stasiun televisi yang memiliki *rating* tinggi lebih berkompeten [5]. Pada penelitian tersebut menggunakan jenis pembelajaran *unsupervised learning*.

Indonesia Lawyers Club (ILC) merupakan *talk show* di tvOne yang menampilkan dialog mengenai masalah hukum dan kriminalitas di Indonesia [6]. Pengelolaan acara yang cukup baik, membuat program ini berhasil mendapatkan peringkat ke-2 [7]. Masih ada satu *talk show* lain yang perlu dihadapi untuk menduduki *rating* nomer 1.

Pada penelitian ini membangun suatu sistem untuk segmentasi opini masyarakat di media sosial twitter menggunakan K-Medoids *Cluster*ing. Perbedaan dengan penelitian segmentasi dengan yang lain, pada penelitian ini proses klasterisasi opini masyarakat akan menggunakan proses pelabelan. Selain itu, data tweet yang sudah diberi label akan digunakan pada tahap prediksi label pada data Tweet episode tayangan ulang.

Analisis terhadap opini masyarakat berhasil menghasilkan informasi yang diperlukan [8]. Sifat "bebas" dari media sosial membuat setiap orang cenderung mengekspresikan pandangannya dalam bentuk komentar [9][10]. Sebuah tinjauan literatur mengidentifikasi 13 studi yang menerapkan metode *clustering* yang berbeda. Hasilnya menunjukkan bahwa penggunaan jenis pembelajaran *unsupervised* dalam menambang data media sosial memiliki beberapa kelemahan [11]. Sebuah penelitian untuk mencoba menekan biaya pada proses klasterisasi mencoba menerapkan proses pelabelan dan berhasil menekan biaya 50-60% [12][13]. Proses ini juga berhasil dilakukan pada algoritma K-Medoids pada metode *active learning* [14][15]. Melihat dari penelitian tersebut, pada penelitian ini akan menggabungkan pembelajaran *supervised learning* dan *unsupervised learning* diharapkan dapat membantu menentukan label pada data Tweet yang akan di olah.

Data Tweet kemudian akan melalui tahap *preprocessing* dan ditransformasi menggunakan perhitungan *term frequency*. Data Tweet yang berbentuk numerik akan diklasterisasi menggunakan K-Medoids *Cluster*ing dengan perhitungan Cosine Similarity. K-Medoids adalah sebuah algoritma *partitional clustering* yang memiliki tujuan memecah data set ke dalam beberapa kelompok. Algoritma K-Medoids dapat mengurangi data baru yang di luar segmen agar masuk ke pusat klaster [16]. Data setiap *cluster* akan di prediksi labelnya berdasarkan mayoritas label yang ada.

2. Metode

2.1. *Data Mining*

Data mining adalah proses menemukan informasi bermakna melalui pola korelasi baru dan trend dengan memilah-milah sejumlah besar data yang disimpan dalam repositori dengan menggunakan teknik pengenalan pola serta teknik statistik dan matematika. Tahapan dalam data *mining* yaitu data *selection*, data *cleaning*, transformasi, data *mining* dan interpretasi.

2.2. Media Sosial Twitter

Media sosial twitter adalah layanan bagi teman, keluarga, dan teman sekerja untuk berkomunikasi dan tetap terhubung melalui pertukaran pesan. Selain dimanfaatkan sebagai sarana berkomunikasi, Tweet juga dapat digunakan dalam tahap analisis [9]. Media sosial twitter cukup baik apabila menjadi objek dalam menganalisis suatu teks.

Teknik analisis dapat menjadi alat yang terbukti dapat mengekstrak informasi. Setiap informasi diperoleh dari berbagai ulasan atau Tweet yang diunggah oleh pengguna twitter. Media sosial twitter cukup baik apabila menjadi objek dalam menganalisis suatu teks [9][10]. Teksnya yang terbatas 140 karakter, membuat informasi yang disampaikan oleh masyarakat lebih bermakna [17].

2.3. *Preprocessing*

Preprocessing dilakukan agar data siap olah. Tahap *preprocessing* yang dilakukan seperti tahapan *text mining* untuk pencarian informasi, klasifikasi teks dan pengelompokan teks [18]. Dalam prosesnya, menggunakan 4 tahap yaitu *case folding*, *tokenizing*, *filtering* dan *stemming*.

Tahap *case folding* merupakan tahap mengubah menjadi jenis yang sama bisa huruf besar maupun huruf kecil dan menghilangkan notasi selain huruf. Tahap *tokenizing* merupakan tahap pemotongan kalimat menjadi kata. Tahap *filtering* merupakan tahap pengambilan data dan penghapusan kata yang tidak digunakan. Terakhir adalah tahap *stemming* yaitu pengelompokan kata-kata lain yang memiliki kata dasar yang serupa.

2.4. Transformasi Data

Tahap transformasi dilakukan dengan *term frequency*. *Term Frequency* (TF) adalah frekuensi dari kemunculan

sebuah term dalam dokumen yang bersangkutan. Semakin besar jumlah kemunculan suatu kata dalam dokumen, semakin besar pula bobotnya atau akan memberikan nilai kesesuaian yang semakin besar [18].

2.5. *Clustering*

Clustering adalah metode pengelompokan data ke dalam kelompok yang berbeda, sehingga data pada masing-masing kelompok memiliki kecenderungan dan pola yang sama. K-Medoids adalah sebuah algoritma yang merepresentasikan *cluster* yang dibentuk menggunakan titik pusat yang berasal dari anggota *cluster*. Tahapannya yaitu:

a. Inisialisasi medoids awal. Untuk menentukan nilai k yang optimal, dapat menggunakan teknik Elbow [11].
b. Perhitungan jarak awal.
c. Aplikasikan setiap data (objek) ke cluster terdekat menggunakan persamaan Cosine Distance dalam menghitung jarak.
d. Memilih acak non-medoids sebagai kandidat medoids baru.
e. Perhitungan jarak medoids lama dengan medoids baru. Jika selisih lebih besar dari 0, mengulang ke tahap a.

Persamaan yang digunakan dalam menghitung jarak yaitu Cosine Distance yang dapat dilihat pada persamaan 1 dan Cosine Similarity pada persamaan 2.

$$Cosine\ Distance = 1 - Similarity \quad (1)$$

$$Similarity\ (x,y) = \cos\theta \frac{\sum_{i=1}^{n} x_i y_i}{\sqrt{\sum_{i=1}^{n} x_i^2}\sqrt{\sum_{i=1}^{n} y_i^2}} \quad (2)$$

dengan x & y adalah dokumen yang dibandingkan, $\sum_{i=1}^{n} x_i y_i$ adalah jumlah bobot kata dokumen x terhadap dokumen y, $\sqrt{\sum_{i=1}^{n} x_i^2}$ adalah akar dari jumlah bobot dokumen x dan $\sqrt{\sum_{i=1}^{n} y_i^2}$ adalah akar dari jumlah bobot dokumen y. Hasil *cluster* akan diuji menggunakan metode Silhouette Coefficient. Silhouette Coefficient merupakan tahapan dimana hasil dari perhitungan yang dilakukan oleh K–Medoids diuji keakuratannya untuk melihat seberapa akurat pengelompokan yang dilakukan oleh metode K–Medoids [19]. Perhitungan menggunakan persamaan 3.

$$S = \frac{b-a}{\max(a,b)} \quad (3)$$

dengan a adalah jarak rata–rata antara medoid dengan obyek di dalam *cluster* dan b adalah jarak rata–rata antara medoid dengan obyek di luar *cluster*. Perhitungan jarak dalam *cluster* dan perhitungan jarak luar *cluster* menggunakan Euclidean Distance dengan persamaan 4.

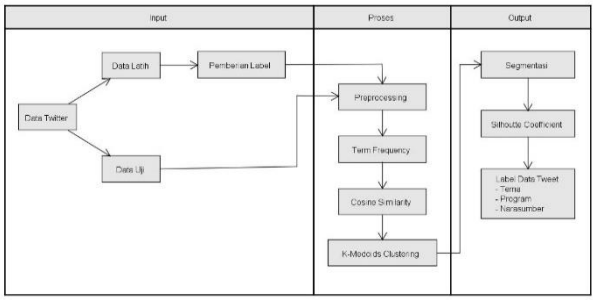
$$d\ (x,y) = \sum \sqrt{(x_i - y_i)^2}\ \ i = 1\ ;\ 1, 2, 3, \ldots n \quad (4)$$

dengan d adalah jarak, x dan y adalah nilai variabel *centroid*.

3. Hasil dan Pembahasan

Sistem dibangun melalui tiga tahapan yaitu *input*, *process* dan *output*. Pada *input* terdapat data twitter yang dibagi menjadi dua yaitu data latih dan data uji. Data uji ini muncul karena metode yang dipakai adalah gabungan dari *supervised* dan *unsupervised* seperti yang sudah disampaikan pada latar belakang.

Pada tahapan input terdapat dua jenis proses yang dijelaskan pada proses data percobaan. Pada tahapan proses dan output terdapat beberapa tahapan yang dimulai dari *preprocessing* hingga mendapatkan label. Ilustrasi dari sistem segmentasi media sosial twitter dapat dilihat pada Gambar 1.



Gambar 1 Ilustrasi Penelitian

3.1. Data Percobaan

Proses *input* menggunakan data dari media sosial twitter dari setiap episode ILC. Tweet yang digunakan merupakan dengan tahun *posting* 2020. Data Tweet yang digunakan akan dibagi menjadi dua jenis yaitu data Tweet saat tayangan pertama dan data Tweet saat tayangan ulang. Data latih ditulis menggunakan format CSV seperti Gambar 2 dan data uji seperti Gambar 3.

	A
1	Jiwasraya kapan di angkat di ILC Bos ???? Takut sama bos besar ya ???
2	sejak menkumham nya si yasina tololi, hukum di Indonesia kacau balau
3	No Rocky No Party ..
4	Kapan JIWASRAYA Bang @karniilyas
5	please undnag yasona please banget bag :(
6	Muke gile... hidup gak ketengkep
20	Minggu depan: simsalabim Lapindo. Keren kayaknya ....
21	Tangkaaap lurah, camat, sopir, kuasa hukum. Sampe keatas
22	lurah nya harus jd narsum jg biar kebobrokan rezim ini terbongkar
23	Jiwasraya kapan dibahas datuk @karniilyas ?
24	Sudah lama saya menanti Om Rocky hadir di @ILCtv1 Namun sampai saat ini penantian itu belum kunjung tiba.!
25	Kapan Om RG diundang lagi ke ILC?

Gambar 2 Data Tweet

Knapa tidak di tangkap kalau tau ada,,
Klo ILC Netral,angkat tuh kasus Jiwasraya dan Lapindo!!! Ada nama AB sempat mencuat!!!!Berani?????
Kapan ILC kasih judul Bakrie dalam pusaran jiwa sraya
ILC yg benar yah spt ini, narasumbernya berbobot pakar hukum, profesor dan pengacara bukan buzzer DO.
Mengapa Rocky Gerung sudah jarang tampil di ILC bang Karni...

Gambar 3 Data Uji

3.2. Pemberian Label

Pada tahap *input* terdapat proses pemberian label pada Tweet. Proses klasterisasi merupakan salah satu jenis pembelajaran *unsupervised*. Namun, pada tahap ini menggunakan pelabelan agar lebih optimal dalam hasil

segmentasi. Terdapat tiga jenis label yang digunakan. Penentuan label dilakukan oleh pihak yang berwenang pada perusahaan. Data yang diberi label yaitu data Tweet saat tayangan pertama (bukan siaran ulang). Data dapat dilihat pada Gambar 4.

1	i»¿Jiwasraya kapan di angkat di ILC Bos ???? Takut sama bos besar ya ???	program
2	sejak menkumham nya si yasina tololi, hukum di Indonesia kacau balau	tema
3	No Rocky No Party ..	narasumber
4	Kapan JIWASRAYA Bang @karniilyas	program
5	[illegible]	narasumber
22	lurah nya harus jd narsum jg biar kebobrokan [illegible]	narasumber
23	Jiwasraya kapan dibahas datuk @karniilyas ?	program
24	Sudah lama saya menanti Om Rocky hadir di @ILCtv1 Namun sampai saat ini penantian itu belum kunjung tiba.!	narasumber
25	Kapan Om RG diundang lagi ke ILC?	narasumber

Gambar 4 Data dengan Label

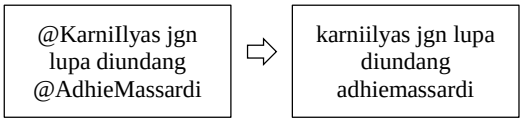
Pada tahap selanjutnya proses dilakukan secara komputasi dengan komputer secara terotomasi. Tahap yang dilakukan secara otomasi yaitu *preprocessing*, transformasi, penentuan *cluster* hingga tahap *output* saat pengujian dan identifikasi.

2.3. Preprocessing

Tahap ketiga yaitu *preprocessing* pada teks. Tahap ketiga ini termasuk tahapan *text mining*. Terdapat 4 tahap yang digunakan dalam *preprocessing* yaitu:

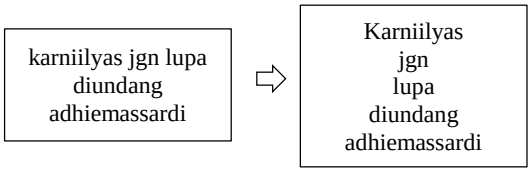
a. Case Folding

Pada tahap pertama ini disebut case folding. Pada prosesnya data diubah semuanya menjadi huruf kecil. Selain itu, karakter selain huruf 'a' sampai dengan 'z' akan dihilangkan dianggap sebagai *delimiter*. *Delimiter* mencakup seluruh karakter selain notasi ASCII akan dihapus.



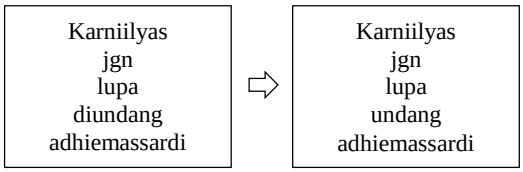
b. Tokenizing

Tahap kedua merupakan tahap pemotongan kalimat. Tahap ini secara garis besar memecah sekumpulan karakter dalam suatu teks. Karakter yang menjadi pembahas seperti spasi, tanda baca atau lainnya yang memiliki fungsi sebagai jeda.



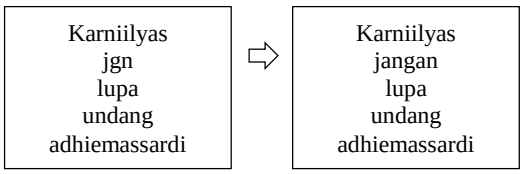
c. Filtering

Tahap ketiga merupakan pengambilan kata dari hasil tahap sebelumnya. Salah satu penerapannya dengan menghilangkan kata-kata yang dianggap *stopword*. Jenis *stopword* seperti kata sambung, kata imbuhan dan kata lain yang memiliki fungsi yang sama. *Stopword* tambahan dibuat disesuaikan dengan kasus. Kata yang terlalu sering disebutkan akan dihapus. Kata tersebut contohnya, ILC.



d. Stemming

Tahap terakhir diperlukan untuk memperkecil jumlah indeks yang berbeda dari suatu dokumen. Tahap ini melakukan pengelompokan kata-kata lain yang memiliki kata dasar yang serupa namun memiliki bentuk yang berbeda. Tahap stemming juga memproses kata yang ditulis tidak baku, salah ketik dan bahasa selain Bahasa Indonesia. Kata yang memiliki makna yang sama atau sinonim akan dihitung kata yang berbeda.



Hasil preprocessing Tweet dapat dilihat pada Gambar 5.

1	jiwasraya angkat bos takut bos
2	menkumham yasonnalaoly hukum indonesia kacau balau
3	rockygerung
4	jiwasraya
21	tangkap lurah camat sopir kuasa hukum
22	lurah narasumber rusak zaman bongkar
23	jiwasraya bahas
24	rockygerung hadir kunjung
25	rockygerung undang

Gambar 5 Hasil Preprocessing

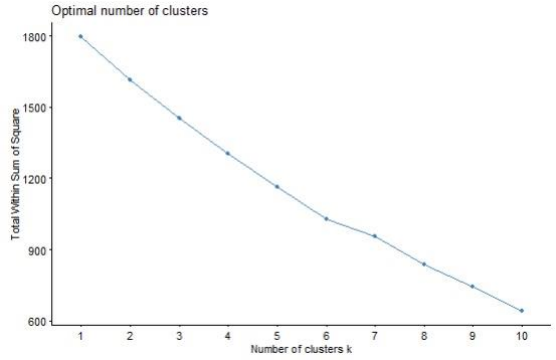
2.4. Transformasi menggunakan TF (Term Frequency)

Transformasi data merupakan proses perhitungan bobot pada teks Tweet yang sudah diolah. Dalam penentuan klaster, data perlu dihitung dengan rumus. Maka Tweet perlu ditransformasikan menjadi data numerik. Salah satu prosesnya data menggunakan akan dihitung berdasarkan *term frequency* (TF).

2.5. Penentuan *Cluster* menggunakan K-Medoids

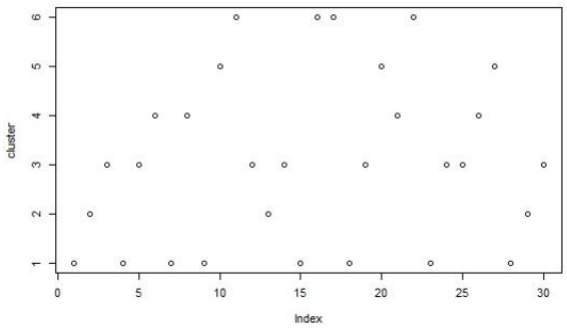
Pada tahap kelima yaitu proses penentuan *cluster* menggunakan Algoritma K-Medoids. Data set yang sudah ditransformasi dihitung dengan rumus k-medoids hingga mendapatkan medoid yang tetap. Syarat mendapatkan medoid tetap adalah apabila selisih antara

total jarak dan total jarak pada medoid baru kurang dari 0. Nilai k ideal ditentukan oleh metode Elbow. Metode Elbow merupakan suatu metode untuk mengetahui informasi dalam menentukan jumlah *cluster* terbaik. Hasil dapat dilihat pada Gambar 6.



Gambar 6 Nilai k dengan Metode Elbow

Grafik diatas menunjukkan siku pada angka 6. Nilai k ideal yang digunakan saat perhitungan klaster dengan K-Medoids adalah 6. Hasil *cluster* dapat dilihat pada Gambar 7.



Gambar 7 Grafik Cluster

Pada gambar 7 diatas sumbu x menunjukan nomer Tweet dan sumbu y menunjukkan letak *cluster*.

2.6. Identifikasi Segmen

Tahap keenam yaitu identifikasi segmen. Tahap ini akan menunjukkan hasil klaster dan label data Tweet. Kesimpulan dapat diambil dari mayoritas label yang muncul pada setiap klaster yang terbentuk. Data Tweet pada episode tayangan ulang akan memiliki label berdasarkan mayoritas label yang terbentuk pada klasternya. Hasil klaster dapat dilihat pada Tabel 1.

Tabel 1 Hasil Klaster

Klaster	Jumlah Data	Mayoritas Label
C1	8 tweet {1, 4, 7, 9, 15, 18, 23, 28}	Program
C2	3 tweet {2, 13, 29}	Tema, Program, Narasumber
C3	8 tweet {3, 5, 12, 14, 19, 24, 25, 30}	Narasumber
C4	4 tweet {6, 8, 21, 26}	Tema
C5	3 tweet {10, 20, 27}	Program
C6	4 tweet {11, 16, 17, 22}	Tema, Narasumber

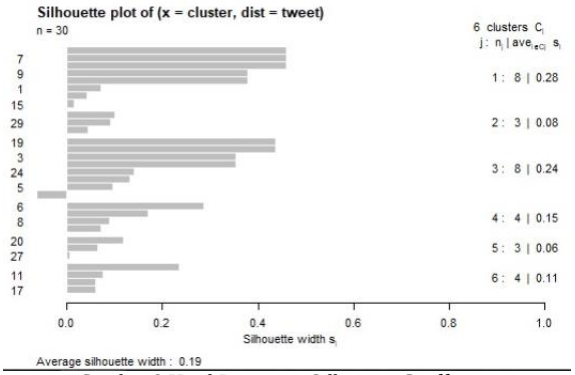
Tweet yang akan dicari labelnya adalah Tweet 26, 27, 28, 29 dan 30. Berdasarkan Tabel 1, maka dapat disimpulkan label dari Tweet yang dicari sesuai mayoritas labelnya. Hasil klasifikasi label dapat dilihat pada tabel 2.

Tabel 2 Hasil Prediksi Label

Tweet	Klaster	Label
26	C4	Tema
27	C5	Program
28	C1	Program
29	C2	Tema, Program, Narasumber
30	C3	Narasumber

2.7. Pengujian

Tahap terakhir yaitu pengujian terhadap sistem yang telah dibangun. Pengujian dilakukan terhadap data untuk uji mutu *cluster* dan uji akurasi hasil label. Pengujian mutu pada klaster yang terbentuk menggunakan metode Silhouette Coefficient. Hasil uji mutu dapat dilihat pada Gambar 8.



Gambar 8 Hasil Pengujian Silhouette Coefficient

Rata-rata dari 6 *cluster* yang terbentuk mendapat mutu 0.19 yang artinya *cluster* yang terbentuk memiliki struktur yang lemah. Pengujian akurasi label yang dapat dilihat pada Tabel 3.

Tabel 3 Hasil Pengujian Akurasi

Pengujian ke-	Hasil
1	79%
2	84%
3	79%
4	60%
5	98%

Berdasarkan hasil dari kelima pengujian mendapatkan rata-rata 80%.

4. Kesimpulan

K-Medoids Clustering sebagai salah satu algoritma clustering berhasil mengklasifikasi data Tweet. Data uji berupa Tweet episode tayangan ulang sebanyak 5 Tweet dan data latih berupa data Tweet pada siaran perdana sebanyak 25 Tweet menghasilkan 6 cluster. Hasil pengujian mutu cluster terbilang memiliki struktur yang lemah karena hanya memiliki nilai 0.19. Hal ini terjadi karena data set yang di transformasi dengan metode term frequency. Kata yang memiliki makna yang sama akan dianggap berbeda oleh metode ini sehingga hasil transformasi kurang representatif. Berbeda dengan pengujian akurasi label, sistem yang dibangun dari kelima pengujian berhasil mendapatkan poin 80% dan dinyatakan berhasil menunjukan opini masyarakat di Twitter terhadap program.

Ucapan Terimakasih

Ucapan terimakasih disampaikan kepada pihak tvOne atas *sample* data twitter yang digunakan sebagai bahan uji coba, khususnya *reply* Tweet dari akun @ILCtv1.

Daftar Rujukan

[1] L. G. B. Ruiz, M. C. Pegalajar, R. Arcucci, and M. Molina-Solana, "A Time-Series Clustering Methodology for Knowledge Extraction in Energy Consumption Data," *Expert Syst. Appl.*, vol. 160, p. 113731, 2020, doi: 10.1016/j.eswa.2020.113731.

[2] P. Arora, Deepali, and S. Varshney, "Analysis of K-Means and K-Medoids Algorithm for Big Data," *Procedia Comput. Sci.*, vol. 78, pp. 507–512, 2016, doi: 10.1016/j.procs.2016.02.095.

[3] S. S. Li, "Lifestyles, Technology Clustering, and the Adoption of Over-the-top Television and Internet Protocol Television in Taiwan," *Int. J. Commun.*, vol. 14, pp. 2017–2035, 2020.

[4] X. Kui *et al.*, "TVseer: A Visual Analytics System for Television Ratings," *Vis. Informatics*, vol. 4, no. 3, pp. 1–11, 2020, doi: 10.1016/j.visinf.2020.06.001.

[5] M. A. Pribadi, M. G. Yoedtadi, and K. H. Siswoko, "Perspektif Praktisi Televisi Indonesia terhadap Konvergensi Televisi dan Internet dalam Persaingan Penyajian Informasi di Internet," *J. Muara Ilmu Sos. Humaniora, dan Seni*, vol. 1, no. 1, p. 319, 2017, doi: 10.24912/jmishumsen.v1i1.372.

[6] Viva, "tvOne Corporate Website," *tvOne*. [Online]. Available: https://tvonenews.tv/. [Accessed: 19-Oct-2019].

[7] Y. Darwis, "Hasil Survei Indeks Kualitas Program Siaran Televisi," Komisi Penyiaran Indonesia Pusat, Jakarta, 2018.

[8] M. D. Devika, C. Sunitha, and A. Ganesh, "Sentiment Analysis: A Comparative Study on Different Approaches," *Procedia Comput. Sci.*, vol. 87, pp. 44–49, 2016, doi: 10.1016/j.procs.2016.05.124.

[9] C. J. Hutto and E. Gilbert, "VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text," in *International AAAI Conference on Weblogs and Social Media*, 2014, pp. 216–225, doi: 10.1210/en.2011-1066.

[10] R. Ahuja, A. Chug, S. Kohli, S. Gupta, and P. Ahuja, "The Impact of Features Extraction on the Sentiment Analysis," *Procedia Comput. Sci.*, vol. 152, pp. 341–348, 2019, doi: 10.1016/j.procs.2019.05.008.

[11] M. Guftar, S. H. Ali, A. A. Raja, and U. Qamar, "A Novel Framework for Classification of Syncope Disease using K-Means Clustering Algorithm," in *SAI Intelligent Systems Conference*, 2015, pp. 127–132, doi: 10.1109/IntelliSys.2015.7361135.

[12] Z. Shuyang, T. Heittola, and T. Virtanen, "Active Learning for Sound Event Classification by Clustering Unlabeled Data," in *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, 2017, pp. 751–755, doi: 10.1109/ICASSP.2017.7952256.

[13] M. Darnstadt, H. Meutzner, and D. Kolossa, "Reducing the Cost of Breaking Audio CAPTCHAs by Active and Semi-supervised Learning," *Proc. - 2014 13th Int. Conf. Mach. Learn. Appl. ICMLA 2014*, pp. 67–73, 2014, doi: 10.1109/ICMLA.2014.16.

[14] Z. Shuyang, T. Heittola, and T. Virtanen, "An Active Learning Method Using Clustering and Committee-Based Sample Selection for Sound Event Classification," *16th Int. Work. Acoust. Signal Enhanc. IWAENC 2018 - Proc.*, pp. 116–120, 2018, doi: 10.1109/IWAENC.2018.8521336.

[15] W. Ji, R. Wang, and J. Ma, "Dictionary-Based Active Learning for Sound Event Classification," *Multimed. Tools Appl.*, vol. 78, no. 3, pp. 3831–3842, 2019, doi: 10.1007/s11042-018-6380-z.

[16] Y. Tan, "An Improved KNN Text Classification Algorithm Based on K-Medoids and Rough Set," in *International Conference on Intelligent Human-Machine Systems and Cybernetics*, 2018, vol. 1, pp. 109–113, doi: 10.1109/IHMSC.2018.00032.

[17] C. N. Dos Santos and M. Gatti, "Deep Convolutional Neural Networks for Sentiment Analysis of Short Texts," in *International Conference on Computational Linguistics*, 2014, pp. 69–78.

[18] S. Vijayarani, M. J. Ilamathi, and M. Nithya, "Preprocessing Techniques for Text Mining -An Overview," *Int. J. Comput. Sci. Commun. Networks*, vol. 5, no. 1, pp. 7–16, 2016.

[19] Y. H. Chrisnanto and G. Abdillah, "Gambaran Umum Kemampuan Akademik Mahasiswa Unjani Dengan Algoritma Partitioning Around Medoids (PAM) Clustering," in *Seminar Nasional Ilmu Pengetahuan dan Teknologi*, 2015, pp. 285–290.