Jurnal Penelitian Inovatif (JUPIN)
Vol. 2, No. 1, April 2022, Hal. 59-66
https://jurnal-id.com/index.php/jupin

DOI: https://doi.org/10.54082/jupin.38
p-ISSN: 2808-148X
e-ISSN: 2808-1366

Analisa Pola Pengguna Youtube Trending Menggunakan Algoritma *Partitioning Around Medoids* dan *Frequent Pattern Growth*

Syarah Seimahuira[*1]

[1]Informatika, Fakultas Teknik, Universitas Nusa Mandiri, Indonesia
Email: [1]syarah.yrs @nusamandiri.ac.id

Abstrak

Seiring dengan perkembangan media digital menjadikan Youtube sebagai salah satu platform media sosial yang memiliki banyak peminatan di dunia. Banyaknya pengguna Youtube yang memanfaatkan platform ini sebagai sarana untuk berkarya dan menghasilkan seperti blog video, video pendidikan, dan berbagai konten lainnya Penelitian ini membahas mengenai pola pengguna youtube trending. Oleh karena itu, *Konten Creator* akan berlomba untuk membuat video agar masuk dalam daftar trending YouTube. Dari data trending tersebut dapat digunakan untuk mengetahui pola pengguna Youtube dan hubungan antara setiap kelompok berdasarkana aktifitas pengguna dalam merespon konten. Penggunaan algoritma Partitioning Around Medoids (PAM) bersamaan dengan algoritma FP-Growth diterapkan dalam melakukan analisa pola pengguna youtube untuk dapat menentukan seberapa besar pengaruh tingkat view terhadap comment, comment count, dislike dan like pada video. Sehingga dengan demikian dapat membantu pengguna youtube dalam mengetahui interaksi dan pengaruhnya terhadap konten yang telah diupload hingga menjadi trending pada youtube. Pengujian dilakukan dengan menerapkan *data mining* dimana pada pengaplikasiannya menggunakan Aplikasi Rapidminer. Dari hasil yang didapatkan berdasarkan pengujian tersebuat adalah ditemukan 12 pola dari sampel dataset Youtube Trending yang dapat digunakan dalam memberikan informasi baru terkait pola masyarakat pengguna Youtube. Hasil tersebut dievaluasi dengan menerapkan *Davies Bouldin Index (DBI)* dimana dilakukan pengujian evaluasi hasil dataset dengan 10 cluster, namun pada kelompok 3 cluster mampu memberikan hasil evaluasi yang mendekati 0 dengan hasil 0,148 sehingga dianggap lebih baik dibandingkan cluster lainnya.

Kata kunci: *Clustering, DBI, FP-Growth, PAM*

Abstract

Along with the development of digital media, Youtube has become one of the most popular social media platforms in the world. Many Youtube users use this platform as a means to create and produce such as video blogs, educational videos, and various other content. Therefore, Content Creators will compete to make videos to be included in the YouTube Trending list. From the trading data, it can be used to find out the pattern of Youtube users and the relationship between each group based on user activity in responding to content. The use of the Partitioning Around Medoids (PAM) algorithm along with the FP-Growth algorithm is applied in analyzing youtube user patterns to find out how much influence view level has on comments, the number of comments, dislikes and likes on videos. So that it can help youtube users find out their interactions and their effects on uploaded content so that it becomes trending on youtube. Testing is done by applying data mining which in its application uses the Rapidminer Application. From the results obtained based on testing, it was found 12 patterns from the Youtube Trending dataset sample that can be used to provide new information related to the patterns of the Youtube user community. These results were evaluated by applying the Davies Bouldin Index (DBI) where an evaluation test of the results of the dataset was carried out with 10 clusters, but cluster 3 was able to provide evaluation results that were close to 0 with a result of 0.148 so it was considered better than other clusters.

Keywords: *Clustering, DBI, FP-Growth, PAM*

1. PENDAHULUAN

Perkembangan Teknologi khususnya media social di Indonesia semakin tidak bisa tebendung dan sudah diprediksi akan semakin meningkat pada tahun 2020. Hal itu tidak lain didorong dengan semakin banyaknya pengguna smartphone yang harganya sangat terjangkau dikalangan masyarakat menengah.

DOI: https://doi.org/10.54082/jupin.38
p-ISSN: 2808-148X
e-ISSN: 2808-1366

Smartphone memudahkan masyarakat untuk mengakses internet dan juga media sosial sebagai tempat mecari berbagai informasi serta berita yang dapat diperoleh secara mudah dan cepat.

Menurut data Asosiasi Penyelenggara Jasa Internet Indonesia (APJII) Indonesia merupakan negara yang memiliki total populasi sebanyak 264 juta jiwa penduduk, ada sebanyak 171,17 juta jiwa atau sekitar 64,8 persen yang sudah terhubung ke internet (litbang, 2019). Satu media yang paling sering dikunjungi yaitu youtube karena sebagai salah satu platform media sosial yang memiliki banyak peminatan di dunia. Dalam Jangkauan Global menurut website Youtube untuk Pers data analisa penguna youtube setiap bulannya lebih dari 1,9 miliar pengguna login mengakses Youtube, setiap harinya lebih dari satu miliar video dilihat hingga menghasilkan miliyaran kali penayangan (David et al., 2017). Pengguna dapat melakukan pengunggahan dengan berbagai format video dan menampilkan berbagai konten seperti klip TV, film, dan video musik.

Banyaknya pengguna Youtube yang memanfaatkan platform ini sebagai sarana untuk berkarya dan menghasilkan seperti blog video, video pendidikan, dan berbagai konten lainnya. Oleh karena itu, *Konten Creator* akan berlomba untuk membuat video agar masuk dalam daftar Trending YouTube. Dari data trending tersebut dapat digunakan untuk mengetahui pola pengguna Youtube dan hubungan antara setiap kelompok berdasarkan aktifitas pengguna dalam merespon konten berdasarkan comment, comment count, dislike dan like pada video. Sehingga penerapan *data mining* diperlukan guna menemukan pengetahuan dari data Trending Youtube menggunakan *Algoritma Partitioning Around Medoids* (PAM) dan FP-Growth sehingga dengan demikian dapat membantu pengguna youtube dalam mengetahui interaksi dan pengaruhnya terhadap konten yang telah diupload hingga menjadi trending pada youtube. Penelitian dilakukan berdasarkan tinjauan studi terdahulu oleh Saputra, Subhi and Winatama pada tahun 2019 yaitu Implementasi Sentimen Analisis Komentar Channel Video Pelayanan Pemerintah di Youtube Menggunakan Algoritma Naive Bayes dengan hasil Channel KemenPUPR menghasilkan nilai nilai 69.23% dan channel KemenKEU menghasilkan nilai 64.10% dan penelitian oleh Tanesab et al., pada tahun 2017 yang melakukan penelitian berupa sentiment analysis model pada media sosial yaitu Youtube menggunakan algoritma *Support Vector Machine* (SVM) untuk mengetahui klasifikasi pattern-pattern yang tidak masuk dalam kelas atau kategori metode yang menjadi pembelajaran (Tanesab et al., 2017).

2. METODE PENELITIAN

Dalam penelitian ini ada beberapa istilah yang digunakan sebagai pendukung proses penelitian

2.1. Youtube

Youtube adalah salah satu platform atau sebuah layanan dari Google yang memfasilitasi penggunanya untuk menguload berbagai video dan dapat diakses oleh pengguna yang lain dari seluruh dunia secara gratis (Nanuru, 2017). Terlepas dari pengertiannya sebagai alat untuk memperesentasikan atau media penyalur berbagai hasil karya, penulis beranggapan bahwa youtube labih dari itu. Pada saat ini youtube menjadi salah satu media untuk memperoleh penghasilan yang menguntungkan hanya dengan membuat konten yang bisa trending dan banyak ditonton orang.

2.2. PAM

Algoritma *clustering Partitioning Around Medoids* (PAM) adalah algoritma untuk menemukan k cluster dalam *n* objek dengan pertama kali secara arbitrarily menemukan wakil dari objek (*medoid*) untuk tiap-tiap cluster (Wardhani, 2017). Clustering merupakan salah satu teknik analisis data eksploratori yang bertujuan untuk menentukan kelompok atau grup dari sekelompok data. Metode ini dikembangkan dengan menentukkan struktur pengelompokkan diantara objek yang akan dikelompokkan

2.3. Algoritma FP-Growth

Algoritma FP-Growth merupakan singkatan dari *Frequent Pattern Growth* yang merupakan salah satu alternatif algoritma yang dapat diterapkan untuk menentukan himpunan data yang sering muncul

DOI: https://doi.org/10.54082/jupin.38
p-ISSN: 2808-148X
e-ISSN: 2808-1366

atau *frequent itemset* dalam sebuah kumpulan data (Nastuti, 2019). Prinsip kerjanya sama dengan Apriori, namun algoritma ini merupakan perkembangannya.

2.4. Sampel Data-set

Data yang digunakan pada penelitian ini adalah Dataset Trending Youtube yang diambil dari Kaggle Repository pada data *Trending Youtube Video Statistics* (Mitchell, 2019). Dataset Youtube Trending tersebut ada sebanyak 31.349 dengan 16 atribut yang menggunakan rumus Slovin sebagai berikut.

$$n = \frac{N}{1+Ne^2} \quad (1)$$

n : ukuran sampel
N : ukuran populasi
e : kelonggaran ketidaktelitian karena kesalahan

Berdasarkan rumus tersebut maka besarnya jumlah sampel adalah sebagai berikut:

$$n = \frac{31.349}{1 + 31.349\ (0{,}0025)} = \frac{31.349}{1 + 78{,}3728} = 394{,}9604712$$

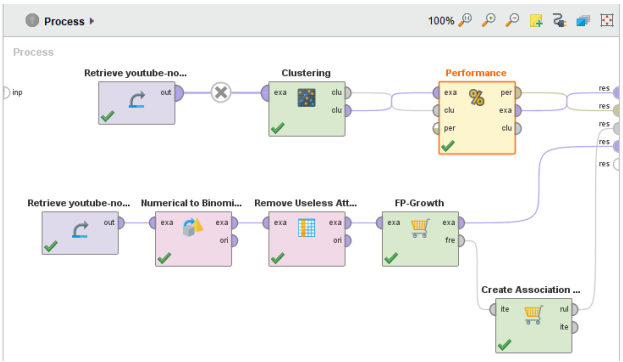
Setelah melakukan pembersihan data Youtube Trending yang digunakan berjumlah 394 buah data terdiri dari 4 atribut. Tipe datanya adalah numerik dan semua data memiliki label kelas.

3. HASIL DAN PEMBAHASAN

Adapun langkah-langkah pembentukan model mining menggunakan alogitma *Algoritma Partitioning Around Medoids* (PAM) dan *FP-Growth* pada *software Rapidminer* antara lain :

a. Memilih file format .csv yang akan dimining
b. Algoritma yang digunakan *Algoritma Partitioning Around Medoids* (PAM) dan *FP-Growth*
c. Mengatur minimal *support*, *confidence* dan rules yang dihasilkan.

Evaluasi data mining akan dilakukan dengan menggunakan dataset youtube. Perhitungan menggunakan algoritma *Algoritma* PAM dan *FP-Growth.* Dataset akan dijalankan melalui model mining yang dibuat pada Rapidminer. Berikut adalah proses pemodelan data menggunakan *Algoritma Partitioning Around Medoids* (PAM) dan *FP-Growth* yang ditampilkan pada Gambar 1.

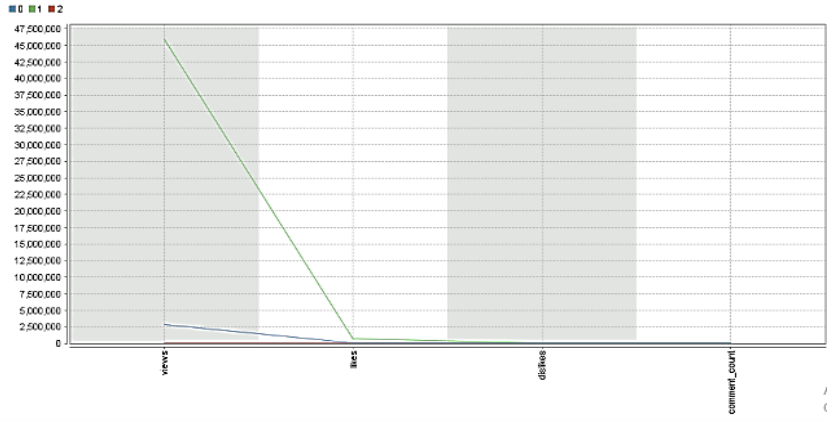


Gambar 1. Proses Pemodelan data menggunakan *Algoritma Partitioning Around Medoids* (PAM) dan *FP-Growth*

3.1. *Performance Cluster Model*

Pada *Performance Cluster Model* menampilkan presentase dari 3 cluster yaitu cluster_0,cluster_1, dan cluster_2 yang ditampilkann dengan garis berwarna hijau, merah dan biru yang ditampilkan pada Gambar 2.

DOI: https://doi.org/10.54082/jupin.38
p-ISSN: 2808-148X
e-ISSN: 2808-1366



Gambar 2. *Performance Vector*

Dari grafik tersebut didapatkan data *centeroid table* yang dijabarkan pada Tabel 1.

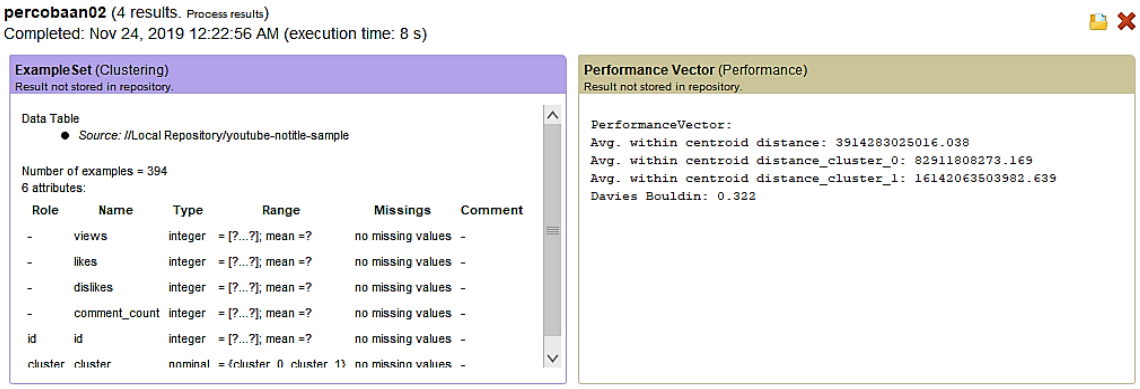
Tabel 1. *Centeroid Table*

Attribute	Cluster_0	Cluster_1	Cluster_2
Views	2866699	45938392	31228
Likes	74831	728696	1159
Dislikes	2734	43049	19
Comment_count	5936	43541	104

Sehingga dapat diketahui banyaknya jumlah penonton Youtube Trending berpengaruh pada like namun tidak tidak memiliki perubahan yang signifikan pada *comment count* dan *dislike*.

3.2. *Example set dan Performance Vector*

Berikut penjelasan detail mengenai proses yang di butuhkan dari total 394 data set dibutuhkan waktu selama 8 detik menggunakan tools untuk melakukan proses mining data. Dapat dilihat pada Gambar 3.



Gambar 3. Example set dan Performance Vector

3.3. *Support* dan *Confidence* menggunakan Rapidminer

Menentukan nilai support dan confidence pada setiap itemset. Pada kasus ini, misalkan diberikan nilai minimum support = 0,9 dan minimum confidence = 1, maka hasilnya yang diperoleh seperti pada Gambar 4.

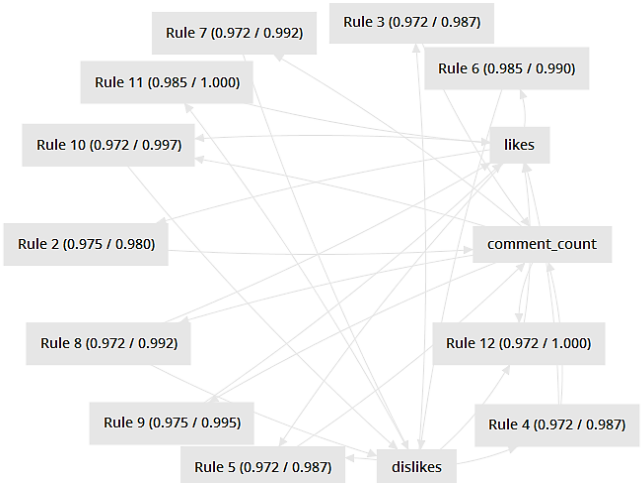
DOI: https://doi.org/10.54082/jupin.38
p-ISSN: 2808-148X
e-ISSN: 2808-1366

...	Premises	Conclusion	Support	Confide...	LaPl...	Gain	p-s	Lift	Convicti...
2	likes	comment_count	0.975	0.980	0.990	-1.015	-0.0...	1.000	0.995
3	dislikes	comment_count	0.972	0.987	0.994	-0.997	0.007	1.008	1.576
4	dislikes	likes, comment_cou...	0.972	0.987	0.994	-0.997	0.012	1.013	1.970
5	likes, dislikes	comment_count	0.972	0.987	0.994	-0.997	0.007	1.008	1.576
6	likes	dislikes	0.985	0.990	0.995	-1.005	0.005	1.005	1.492
7	comment_count	dislikes	0.972	0.992	0.996	-0.987	0.007	1.008	1.959
8	comment_count	likes, dislikes	0.972	0.992	0.996	-0.987	0.007	1.008	1.959
9	comment_count	likes	0.975	0.995	0.997	-0.985	-0.0...	1.000	0.980
10	likes, comment_co...	dislikes	0.972	0.997	0.999	-0.977	0.012	1.013	5.848
11	dislikes	likes	0.985	1	1	-0.985	0.005	1.005	∞
12	dislikes, comment...	likes	0.972	1	1	-0.972	0.005	1.005	∞

Gambar 4. *Support* dan *Confidence* menggunakan Rapidminer

3.4. *Association Rules*

Dengan menerapkan Algoritma *Association Rules* maka diketahui 12 pola premises rule dengan pola yang diperoleh pada grafis rule sampel data youtube pada Gambar 5 dan rule yang dihasilkan pada Gambar 6.



Gambar 5. Grafis Rule Sampel data Youtube

```
Association Rules
[likes] --> [dislikes, comment_count] (confidence: 0.977)
[likes] --> [comment_count] (confidence: 0.980)
[dislikes] --> [comment_count] (confidence: 0.987)
[dislikes] --> [likes, comment_count] (confidence: 0.987)
[likes, dislikes] --> [comment_count] (confidence: 0.987)
[likes] --> [dislikes] (confidence: 0.990)
[comment_count] --> [dislikes] (confidence: 0.992)
[comment_count] --> [likes, dislikes] (confidence: 0.992)
[comment_count] --> [likes] (confidence: 0.995)
[likes, comment_count] --> [dislikes] (confidence: 0.997)
[dislikes] --> [likes] (confidence: 1.000)
[dislikes, comment_count] --> [likes] (confidence: 1.000)
```

Gambar 6. Rules yang Dihasilkan

DOI: https://doi.org/10.54082/jupin.38
p-ISSN: 2808-148X
e-ISSN: 2808-1366

Sehingga dapat dijabarkan sebagai berikut:

a. Jika video mendapatkan like maka dislike dan coment-count mempengaruhi confidence =0,977
b. Jika video mendapatkan like maka comment-count mempengaruhi confidence = 0,980
c. Jika video mendapatkan dislike maka coment-count mempengaruhi confidence = 0,987
d. Jika video mendapatkan dislike maka likes dan coment-count mempengaruhi confidence= 0,987
e. Jika video mendapatkan like atau dislike maka comment_cont mempengaruhi confidence= 0,987
f. Jika video mendapatkan like maka dilikes mempengaruhi confidence= 0,990
g. Jika video mendapatkan coment-count maka dilikes mempengaruhi confidence= 0,992
h. Jika video mendapatkan coment-count maka like atau dilikes mempengaruhi confidence= 0,992
i. Jika video mendapatkan coment-count maka like mempengaruhi confidence= 0,995
j. Jika video mendapatkan like dan coment-count maka dislike mempengaruhi confidence= 0,997
k. Jika video mendapatkan dislike maka like mempengaruhi confidence= 1,000
l. Jika video mendapatkan dislike dan commernt_count maka like mempengaruhi confidence= 1,000

3.5. *Davis Bouldin*

Dalam melakukan evaluasi penelitian pada nilai cluster terbaik yang diperoleh dengan melakukan percobaan beberapa kali pada *Davies Bouldin Index* (DBI) maka didapatkan hasil seperti yang tertera pada Tabel 2.

Tabel 2. *Davies Bouldin Index*

Nilai K	Hasil
1	0,322
2	0,148
3	**0,255**
4	0,398
5	0,952
6	0,216
7	0,241
8	0,404
9	0,225
10	0,322

Sehingga dari table diatas diperoleh nilai terbaik yaitu dari nilai K = 3 dengan hasil 0,148. Dengan ketentuan semakin nilai yang dihasilkan dari nilai K mendekati angka ‘0’ maka hasil disebut data terbaik.

4. KESIMPULAN

Berdasarkan analisa dan hasil pengujian maka dapat diperoleh kesimpulan bahwa data mining dengan *Algoritma Partitioning Around Medoids* (PAM) dan *FP-Growth* dapat diimplementasikan dengan data set mengenai trending youtube. Penelitian ini dapat membantu memberikan informasi bahwa adanya faktor yang mempengaruhi video di youtube menjadi trending diantaranya comment, comment count, dislike dan like pada video. Pada sampel video trending yang digunakan dalam penelitian diperoleh pola masyarakat dalam menggunakan Youtube bahwa lebih banyak penonton Youtube tidak memberikan respon terhadap video yang ditonton baik berupa like dislike maupun comment. Sehingga jumlah viewers berbanding jauh dengan jumlah respon atau interaksi dari penonton.

DAFTAR PUSTAKA

David, E. R. (Eribka), Sondakh, M. (Mariam), & Harilama, S. (Stefi). (2017). Pengaruh Konten Vlog Dalam Youtube Terhadap Pembentukan Sikap Mahasiswa Ilmu Komunikasi Fakultas Ilmu Sosial Dan Politik Universitas Sam Ratulangi. *Acta Diurna*, *6*(1), 93363. https://www.neliti.com/publications/93363/pengaruh-konten-vlog-dalam-youtube-terhadap-

DOI: https://doi.org/10.54082/jupin.38
p-ISSN: 2808-148X
e-ISSN: 2808-1366

pembentukan-sikap-mahasiswa-ilmu-kom

litbang. (2019). *APJII: Jumlah Pengguna Internet di Indonesia Tembus 171 Juta Jiwa*. Litbang Kemendagri. https://litbang.kemendagri.go.id/website/apjii-jumlah-pengguna-internet-di-indonesia-tembus-171-juta-jiwa/

Mitchell, J. (2019). *Trending YouTube Video Statistics*. Kaggle. http://www.kaggle.com/datasnaek/youtube-new

Nanuru, R. F. (2017). *YOUTUBE: Seni Berwawasan Teknologi Modern*. https://doi.org/10.31219/osf.io/3vshc

Nastuti, A. (2019). Amelia Nastuti 1) , Syaiful Zuhri Harahap 2). *Teknik Data Mining Untuk Penentuan Paket Hemat Sembako Dan Kebutuhan Harian Dengan Menggunakan Algoritma Fp-Growth*, *7*(3), 111–119.

Saputra, P. Y., Subhi, D. H., & Winatama, F. Z. A. (2019). Implementasi Sentimen Analisis Komentar Channel Video Pelayanan Pemerintah Di Youtube Menggunakan Algoritma Naïve Bayes. *Jurnal Informatika Polinema*, *5*(4), 209–213. https://doi.org/10.33795/jip.v5i4.259

Tanesab, F. I., Sembiring, I., & Purnomo, H. D. (2017). Sentiment Analysis Model Based On Youtube Comment Using Support Vector Machine. *International Journal of Computer Science and Software Engineering (IJCSSE)*, *6*(8), 180–185. http://ijcsse.org/published/volume6/issue8/p2-V6I8.pdf

Wardhani, A. K. (2017). Penerapan Algoritma Partitioning Around Medoids Untuk Menentukan Kelompok Penyakit Pasien (Studi Kasus : Puskesmas Kajen Pekalongan). *Jurnal Kajian Ilmu Dan Teknologi (KILAT)*, *6*(1), 6–10.

Jurnal Penelitian Inovatif (JUPIN)
Vol. 2, No. 1, April 2022, Hal. 59-66
https://jurnal-id.com/index.php/jupin
DOI: https://doi.org/10.54082/jupin.38
p-ISSN: 2808-148X
e-ISSN: 2808-1366

Halaman Ini Dikosongkan