

CLUSTERING PENGGUNAAN INTERNET PADA SISWA SEKOLAH DASAR DI KOTA TERNATE TENGAH DENGAN MENGGUNAKAN ALGORITMA K-MEANS

Hesti Sukawati¹, Abdul Mubarak², Firman Tempola³

^{1,2,3} Program Studi Informatika Universitas Khairun

Email: ¹ hestisukma18@gmail.com, ² amuba@unkhair.ac.id, ³ firman.tempola@unkhair.ac.id

(Naskah masuk: 24 November 2023, diterima untuk diterbitkan: 5 Desember 2023)

Abstrak

Penelitian ini bertujuan untuk mengelompokan (clustering) penggunaan internet siswa Sekolah Dasar di Kota Ternate Tengah dengan menggunakan algoritma k-means dengan menentukan Sekolah-sekolah yang aktif menggunakan internet. Maka dari itu dengan adanya penelitian ini dengan menggunakan algoritma k-means clustering dapat sedikit membantu dalam memberikan informasi kepada para pihak Sekolah, Orang tua dan Masyarakat. Data yang diambil yaitu data penggunaan internet pada siswa Sekolah Dasar kelas v (lima) data yang diambil ditentukan berdasarkan kriteria yaitu sosial media, youtube, dan game. Metode evaluasi cluster menggunakan metode DBI (Davies Bouldin Index) untuk mengetahui seberapa baik cluster yang dipakai pada penelitian, penulis sudah melakukan perhitungan cluster yaitu 3 cluster untuk menentukan hasil hitung hasilnya, hasil dari DBI yaitu 0,85. Hasil penelitian menggunakan 20 data penggunaan internet. Mana ada 4 data yang masuk pada cluster tiga (C3), 8 data yang masuk pada cluster dua (C2), dan 8 data yang masuk pada cluster satu (C1).

Kata kunci: *Internet, Sekolah, Cluster, K-Means, Davies Bouldin Index*

CLUSTERING INTERNET USE IN PRIMARY SCHOOL STUDENTS IN CENTRAL TERNATE CITY USING THE K-MEANS ALGORITHM

Abstract

This research aims to classify (clustering) internet usage on Elementary Students in the Central Ternate City by using K-Means Clustering Algorithm to determine which school are actively using the internet. So, with this research by using K-Means Clustering Algorithm can be a little helpful for give information to the parties of the school, parents and society. The data taken from the internet usage in Elementary School Students in Class Five, the data taken based on criteria namely social media, youtube and game. The cluster evaluation method using DBI Method to find out how well the cluster is used in the research, the author has done a cluster calculation that are 3 clusters to determine the results of the calculated result, the result of DBI are 0.85. the result of this research used 20 internet usage data. there are 4 data entered in cluster three (C3), 8 data are based on cluster two (C2), and 8 data are based on cluster one (C1).

Keywords: *Internet, School, Cluster, K-Means, Davies Bouldin Index*

1. PENDAHULUAN

Saat ini penggunaan internet sudah menjadi bagian hidup bagi masyarakat global, tidak hanya terbatas bagi orang dewasa saja namun juga anak-anak, umumnya mereka sekarang ini sudah menggunakan internet. Dimana fasilitas teknologi yang memudahkan semua orang untuk lebih mudah mengakses konten-konten melalui software – software sehingga apapun dapat dicari, ditemui, dan di pelajari oleh semua orang kapanpun dimanapun sesuai dengan kebutuhan dan keperluannya masing-masing.

Kemajuan teknologi komputer, smartphone dan internet dimanfaatkan pada berbagai bidang misalnya mengerjakan tugas sekolah, belajar, mengatur keuangan keluarga, mendengarkan musik, menonton video, dan menikmati permainan.

Namun, di balik kemudahan yang ditawarkannya, internet juga dapat membawa sisi negatif bagi penggunaannya. Kekhawatiran yang muncul dari tidak terbandungnya akses internet adalah dunia anak-anak dan remaja, dimana mereka masih sangat labil bahkan belum mampu secara tegas membedakan mana yang baik dan buruk untuk diri mereka sendiri. Maka dari itu di perlukannya

pengelompokan tingkat penggunaan internet pada siswa sekolah dasar agar dapat membantu para orang tua untuk memantau tingkat penggunaan internet anak. (Emir dkk, 2016)

Beberapa dampak negatif yang dapat ditimbulkan dari penggunaan internet adalah adanya konten atau gambar yang tidak pantas, chatting seksual, baik dari media sosial maupun email, kekerasan atau pelecehan online, terbukanya informasi pribadi dan adanya pencurian data dengan cara sharing atau lainnya, situs-situs yang dibuka mengandung spyware, virus dan malicious software, penipuan,

komersialisasi yang berlebihan, adanya keinginan untuk melakukan pembajakan software, musik, atau video. Penelitian tentang penggunaan internet di kota di kota Ternate pernah diteliti oleh (Kapita, 2022) dengan menggunakan metode Apriori dengan obyek penelitian yaitu pada mahasiswa.

Dari hasil penelitian yang dilakukan oleh APJII (Asosiasi Penyelenggara Jasa Internet Indonesia) yang bekerja sama dengan LPI (Lembaga Poling Indonesia) dan penelitian sebelumnya oleh beberapa peneliti maka penulis mengangkat tiga kriteria konten yang sering di kunjungi anak-anak yaitu YouTube, Sosial Media dan Game Online untuk melihat tingkat penggunaan internet anak-anak Sekolah Dasar di Kota Ternate.

2. METODE PENELITIAN

2.1 Data Mining

Data mining adalah proses menyaring data yang sangat besar yang tersimpan dalam penyimpanan dan menemukan hubungan, pola, dan tren baru yang penting dengan menggunakan teknik pengenalan pola seperti matematika dan statistik. Data mining dibagi menjadi empat model: klasifikasi, prediksi, clustering, dan asosiasi.

2.2 K-Means clustering

Pada dasarnya, clustering adalah cara untuk mencari dan mengelompokkan data yang memiliki ciri-ciri yang menyerupai satu sama lain. Clustering adalah salah satu metode data mining yang bersifat tanpa arahan. Ini berarti bahwa metode ini dapat diterapkan tanpa adanya instruksi atau instruktur, dan tanpa adanya output yang diinginkan. Clustering hierarchical dan non-hierarchical adalah dua jenis clustering yang digunakan dalam data mining. Johan (2013)

Metode pengelompokan data yang dikenal sebagai kelompok hierarkis dimulai dengan mengelompokkan dua atau lebih objek yang memiliki kesamaan paling dekat, kemudian melanjutkan ke objek yang memiliki kesamaan kedua, dan seterusnya sampai kelompok membentuk pohon dengan hierarki (tingkatan) yang jelas antar objek, dari yang paling mirip hingga yang paling tidak mirip. Secara logika, setiap objek hanya akan membentuk sebuah cluster pada akhirnya. Prosedur hierarki biasanya lebih jelas dengan dendrogram. Johan (2013)

Algoritma K-Means adalah algoritma yang cukup sederhana untuk mengklasifikasikan atau mengelompokkan sejumlah besar objek yang memiliki karakteristik tertentu ke dalam kelompok (cluster) sebanyak K. Jumlah cluster K sudah ditentukan lebih dahulu. Edmira (2010)

Cluster masing-masing memiliki titik pusat, dan anggota cluster terdekat dipilih berdasarkan jarak dari titik pusat cluster terdekat. Karena penambahan satu anggota pada lokasi yang signifikan akan mengubah lokasi titik pusat cluster dan status keanggotaan harus ditinjau kembali, perubahan keanggotaan kemudian akan mengubah lokasi titik pusat cluster lagi, dan status keanggotaan mungkin berubah lagi, penentuan titik pusat dan keanggotaan cluster kemudian menjadi sulit. Edmira (2010) Langkah-langkah melakukan clustering menggunakan metode K-Means adalah sebagai berikut.

1. Tentukan jumlah cluster k.
2. Inisialisasi k pusat cluster secara acak.
3. Alokasikan semua data ke cluster terdekat.

Berdasarkan persamaan 1, rumus jarak geometris dapat digunakan untuk menentukan seberapa dekat data ke setiap pusat cluster. Jarak antara data dan pusat cluster menentukan seberapa jauh data termasuk dalam cluster.

$$d(i, j) = \sqrt{\sum_{l=1}^p (x_{il} - x_{jl})^2} \quad (1)$$

Keterangan: p adalah jumlah atribut; l adalah 1,2,3,...,p; x_{il} adalah data ke-i pada atribut ke-l; $d(i, j)$ adalah jarak data ke-i terhadap pusat cluster ke-j; dan x_{jl} adalah pusat cluster ke-j pada atribut ke-l.

Berdasarkan keanggotaan baru, hitung kembali pusat kelompok. Persamaan 2 menunjukkan rumus untuk menghitung pusat cluster baru, yang dihitung berdasarkan rata-rata data untuk setiap atribut dari satu cluster.

$$v_j = \frac{\sum_{i=1}^{N_j} \sum_{l=1}^p x_{il}}{N_j} \quad (2)$$

Keterangan: v_j adalah pusat cluster j yang baru; N_j adalah jumlah data yang menjadi anggota cluster j; i adalah 1,2,3,..., N_j ; p adalah jumlah atribut; l adalah 1,2,3,...,p; x_{il} adalah data ke-i pada atribut l.

Hitung jarak antara data dan pusat cluster baru. Proses clustering selesai ketika anggota cluster tidak berpindah dari satu cluster ke cluster lain. Jika anggota cluster berpindah dari satu cluster ke cluster lain, maka kembali ke langkah 3 hingga tidak ada lagi anggota cluster yang berpindah. Evi et al., (2017).

2.3 Davies bouldin index

Salah satu metode untuk validasi kelompok yang dikembangkan oleh DL Davies adalah Davis Bouldin Index (DBI). DBI adalah salah satu pendekatan evaluasi kelompok dalam pendekatan pengelompokan yang didasarkan pada nilai kohesi

dan separasi. Kohesi dalam pengelompokan adalah jumlah kedekatan data terhadap centroid dari cluster yang diikuti, sedangkan separasi adalah jumlah jarak antar centroid cluster. Matrik kohesi cluster ke-i, yang dirumuskan sebagai berikut, dapat dihitung dengan persamaan jumlah persegi dalam cluster (SSW) :

$$SSW_i = \frac{1}{m_i} \sum_{j=i}^{m_i} d(x_j, c_i) \dots \dots \dots (3)$$

Keterangan:

m_i = jumlah data dalam cluster ke-i

c_i = centroid cluster ke-i

$d()$ = jarak setiap data ke centroid yang dihitung menggunakan jarak *Euclidean*

Sum of square between cluster (SSB) merupakan persamaan yang digunakan untuk mengetahui separasi antar cluster yang dihitung menggunakan persamaan:

$$SSB_{i,j} = d(c_i, c_j) \dots \dots \dots (4)$$

Setelah nilai kohesi dan separasi didapat, pengukuran rasio (R_{IJ}) dilakukan untuk mengetahui nilai perbandingan antara cluster I dan cluster J. Kelompok dengan nilai kohesi dan separasi yang paling rendah adalah yang terbaik. Persamaan berikut digunakan untuk menghitung nilai rasio:

$$R_{ij} = \frac{SSW_i + SS_j}{SSB_{i,j}} \dots \dots \dots (5)$$

Nilai rasio yang diperoleh tersebut digunakan untuk mencari nilai *davies bouldin index*

(DBI) dari persamaan berikut :

$$DBI = \frac{1}{k} \sum_{i=1}^k \max_{j \neq i} (R_{i,j}) \dots \dots \dots (4)$$

Dari persamaan tersebut, k merupakan *cluster* yang digunakan. Semakin kecil nilai DBI yang diperoleh (non-negatif ≥ 0), maka semakin baik *cluster* yang diperoleh dari pengelompokan *K-Means* yang digunakan (Alith, 2018).

2.4 Dataset

Dataset yang digunakan dalam penelitian yaitu hasil pengumpulan kuesioner dari 20 sekolah di Kota Ternate. Kemudian penggunaan internet diukur dari penggunaan social media, youtube dan game online.

3. HASIL DAN PEMBAHASAN

3.1 Hasil Pengumpulan Data

Penelitian ini dibuat dengan menggunakan data primer, dengan jumlah populasi sebanyak 20 Sekolah Dasar dan jumlah respondenya adalah 634 siswa kelas 5 dari masing-masing sekolah di kota ternate tengah.

Parameter yang digunakan untuk melakukan pengelompokan ada 3 yaitu sosial media, youtube dan games jumlah data yang akan di gunakan sebanyak 20 data. dapat diolah lebih lanjut. Jika semua butir pertanyaan yang ada dijawab sesuai dengan cara pengisian kuesioner, maka data kuesioner tersebut dikatakan layak sehingga dapat diolah lebih lanjut.

Tabel 1. jumlah responden

No	Nama SD	jumlah siswa kls 5 yang hadir	yang aktif bersosmed	Tidak aktif
1	SDN 1 Kota Ternate	42	28	14
2	SDN 10 Kota Ternate	35	20	15
3	SDN 11 Kota Ternate	32	15	17
4	SDN 12 Kota Ternate	30	16	14
5	SDN 13 Kota Ternate	31	10	21
6	SDN 14 Kota Ternate	27	7	20
7	SDN 15 Kota Ternate	30	20	10
8	SDN 16 Kota Ternate	26	22	4
9	SDN 17 Kota Ternate	25	17	8
10	SDN 18 Kota Ternate	22	5	17
11	SDN 19 Kota Ternate	24	15	9
12	SDN 2 Kota Ternate	40	12	28
13	SDN 3 Kota Ternate	30	14	16
14	SDN 4 Kota Ternate	32	16	16
15	SDN 5 Kota Ternate	39	7	32
16	SDN 6 Kota Ternate	37	10	14
17	SDN 7 Kota Ternate	40	11	15
18	SDN 8 Kota Ternate	35	9	17
19	SDN 9 Kota Ternate	28	8	14
20	SD PERTIWI 1	32	10	21

Parameter Tingkat Penggunaan Internet

Pada tahap ini di lakukan perhitungan untuk mencari nilai rata-rata dari masing-masing parameter dengan rumus yang digunakan untuk mencari nilai rata-rat yaitu :

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} \dots \dots \dots (6)$$

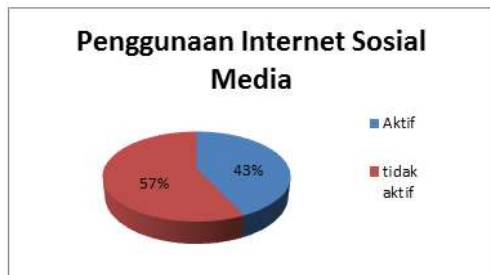
Keterangan :

$\bar{x}$ = mean

n = banyaknya data

x_i = nilai data ke-i

Berikut adalah hasil nilai rata - rata dari masing - masing parameter yang telah di hitung berdasarkan masing-masing item dibuat dalam bentuk grafik.



Gambar 1. Grafik Presentasi Penggunaan Internet Sosial Media



Gambar 2. Grafik Presentasi Penggunaan Internet Youtube



Gambar 3. Grafik Presentasi Penggunaan Internet Game Online

Analisis Data Kriteria

Pada tahap ini akan dilakukan perhitungan data kriteria yang dihitung dari jumlah responden yang menjawab sering menggunakan internet atau yang aktif, kemudian dari data tersebut dihitung dengan rumus perhitungan presentase yaitu :

$$P = \frac{f}{n} \times 100\%$$

Keterangan :

P = Presentase

f = Jumlah jawaban

n = Jumlah responden

Dari responden siswa kelas 5 Sekolah Dasar yang menjawab sering atau yang aktif mengakses Sosial Media, Youtube, dan Game Online. Berikut hasil perhitungan presentase siswa kelas 5 tiap-tiap Sekolah Dasar di Kota Ternate Tengah berdasarkan kriteria yang diangkat yaitu dapat dilihat pada tabel 2.

Tabel 2. Data Penggunaan Internet pada Siswa Sekolah Dasar

No	Nama Sekolah Dasar	Sosial Media	Youtube	Games
1	SD Negeri 1 Kota Ternate	0,667	0,714	0,762
2	SD Negeri 10 Kota Ternate	0,571	0,714	0,857
3	SD Negeri 11 Kota Ternate	0,469	0,906	0,938
4	SD Negeri 12 Kota Ternate	0,533	0,667	0,933
5	SD Negeri 13 Kota Ternate	0,323	0,323	0,968
6	SD Negeri 14 Kota Ternate	0,250	0,321	0,857
7	SD Negeri 15 Kota Ternate	0,667	0,667	0,833
8	SD Negeri 16 Kota Ternate	0,846	0,846	0,769
9	SD Negeri 17 Kota Ternate	0,680	0,800	0,960
10	SD Negeri 18 Kota Ternate	0,227	0,455	0,955
11	SD Negeri 19 Kota Ternate	0,750	0,850	0,900
12	SD Negeri 2 Kota Ternate	0,300	0,300	0,525
13	SD Negeri 3 Kota Ternate	0,467	0,467	0,833
14	SD Negeri 4 Kota Ternate	0,533	0,533	0,900
15	SD Negeri 5 Kota Ternate	0,179	0,231	0,821
16	SD Negeri 6 Kota Ternate	0,270	0,270	0,811
17	SD Negeri 7 Kota Ternate	0,275	0,275	0,750
18	SD Negeri 8 Kota Ternate	0,257	0,714	0,857
19	SD Negeri 9 Kota Ternate	0,267	0,933	0,667
20	SD Pertiwi 1	0,313	0,688	0,938

3.2 Hitung Clustering Menggunakan Algoritma K-Means

Pada tahap ini akan dilakukan perhitungan *Clustering Data*, dimana data dari masing-masing kriteria yang telah di hitung akan diklaster menggunakan tahap algoritma *K-Means*. Karena yang dihitung adalah penggunaan internet maka peneliti menentukan 3 kriteria, yaitu : Sosial media, Youtube, dan Game online.

Clustering Penggunaan Internet

Berikut ini erupakan langkah-langkah clustering:

- Menentukan jumlah *cluster*, yaitu dengan mengelompokan data pada Tabel 4.4 menjadi 3 kelompok *Cluster*, dimana C_1 (*Cluster* satu) dikategorikan sebagai kurang aktif, C_2 (*Cluster* dua) dikategorikan sebagai cukup aktif, dan C_3 (*Cluster* tiga) dikategorikan aktif.
- Menentukan *centroid* awal secara acak, Dimana jumlah cluster = 3. Ditentukan:
 $C_1 = 231; 532; 367$.
 $C_2 = 367; 532; 231$.
 $C_3 = 231; 367; 532$.

- c. Hitung jarak setiap data yang ada pada Tabel 3 terhadap setiap pusat *cluster* dengan menggunakan persamaan *Euclidean Distance*:

$$d_{(x,y)} = \sqrt{(xi - yi)^2 + (xi - yi)^2} \dots$$

Jarak data dengan pusat *cluster* $C_1 = 231; 532; 367, C_2 = 367; 532; 231, C_3 = 231; 367; 532$. Berikut hasil perhitungan jarak setiap data terhadap setiap pusat *cluster* pada iterasi pertama yang disajikan pada Tabel 3

Tabel 3. Posisi *Cluster* Pada Iterasi Pertama

No	Nama Sekolah	C1	C2	C3	Posisi Cluster
1	SD Negeri 1 Kota Ternate	685.1 63	685.18 2	685.1 51	C_3
2	SD Negeri 10 Kota Ternate	685.1 45	685.20 1	685.1 1	C_3
3	SD Negeri 11 Kota Ternate	684.9 87	685.08	684.9 79	C_3
4	SD Negeri 12 Kota Ternate	685.1 53	685.23 3	685.0 89	C_3
5	SD Negeri 13 Kota Ternate	685.4 72	685.6	685.3 17	C_3
6	SD Negeri 14 Kota Ternate	685.5 57	685.67 8	685.4 28	C_3
7	SD Negeri 15 Kota Ternate	685.1 62	685.19 4	685.1 22	C_3
8	SD Negeri 16 Kota Ternate	684.9 97	684.98 1	685.0 15	C_2
9	SD Negeri 17 Kota Ternate	684.9 86	685.04 2	684.9 48	C_3
10	SD Negeri 18 Kota Ternate	685.4 09	685.55 3	685.2 88	C_3
11	SD Negeri 19 Kota Ternate	684.9 56	684.98 6	684.9 44	C_3
12	SD Negeri 2 Kota Ternate	685.7 34	685.77 9	685.6 8	C_3
13	SD Negeri 3 Kota Ternate	685.3 84	685.45 7	685.2 96	C_3
14	SD Negeri 4 Kota Ternate	685.2 75	685.34 8	685.1 86	C_3
15	SD Negeri 5 Kota Ternate	685.6 7	685.79 8	685.5 28	C_3
16	SD Negeri 6 Kota Ternate	685.6 15	685.72 2	685.4 84	C_3
17	SD Negeri 7 Kota Ternate	685.6 42	685.73 6	685.5 27	C_3
18	SD Negeri 8 Kota Ternate	685.2 5	685.36 9	685.2 16	C_3
19	SD Negeri 9 Kota Ternate	685.1 79	685.25 8	685.2 43	C_1
20	SD Pertiwi 1	685.2 08	685.33 2	685.1 48	C_3

3.3 Validasi *davies bouldin index*

Pada tahap ini penulis akan menghitung validasi *cluster* menggunakan metode DBI untuk menentukan *cluster* terbaik. DBI merupakan salah satu metode evaluasi *cluster* pada suatu metode pengelompokan yang didasarkan pada nilai kohesi (mencari nilai SSW atau Sum Square Within) dan separasi (mencari nilai SSB atau Sum Square Between). Dalam suatu pengelompokan, kohesi didefinisikan sebagai jumlah dari kedekatan data terhadap *centroid* dari *cluster* yang diikuti. Sedangkan separasi didasarkan pada jarak antar *centroid* dari *clusternya*. Menentukan hasil dari SSW dan SSB dengan rumus pengukuran jarak dengan menggunakan nilai akhir *centroid* atau nilai titik pusat *cluster* pada hasil hitungan algoritma *k-means*.

Setelah nilai SSW dan SSB diperoleh maka lanjut untuk mencari nilai rasio, dari nilai rasio yang didapat maka bisa ditentukan nilai DBInya. Semakin kecil nilai DBI yang didapat (mendekati nilai 0) semakin baik *cluster* yang diperoleh dalam penelitian ini.

Perhitungan Validasi

1. Hitung jarak data dalam *cluster*

- a. *Cluster* pertama memiliki pusat (0,327; 0,810; 0,850). Ada 4 data yaitu data ke-3, data ke-18, data ke-19 dan data ke-20. Sehingga:

$$C_{1,3} = \sqrt{(0,327 - 0,469)^2 + (0,810 - 0,906)^2 + (0,850 - 0,938)^2} = 0,193$$

$$C_{1,18} = \sqrt{(0,327 - 0,257)^2 + (0,810 - 0,714)^2 + (0,850 - 0,857)^2} = 0,119$$

$$C_{1,19} = \sqrt{(0,327 - 0,267)^2 + (0,810 - 0,933)^2 + (0,850 - 0,667)^2} = 0,229$$

$$C_{1,20} = \sqrt{(0,327 - 0,313)^2 + (0,810 - 0,688)^2 + (0,850 - 0,938)^2} = 0,151$$

- b. *Cluster* kedua C_2 memiliki pusat (0,656; 0,724; 0,864). Ada 8 data yaitu data ke-1, data ke-2, data ke-4, data ke-7, data ke-8, data ke-9 data ke-11 dan data ke-14 sehingga:

$$C_{2,1} = \sqrt{(0,656 - 0,667)^2 + (0,724 - 0,714)^2 + (0,864 - 0,762)^2} = 0,103$$

$$C_{2,2} = \sqrt{(0,656 - 0,571)^2 + (0,724 - 0,714)^2 + (0,864 - 0,857)^2} = 0,086$$

$$C_{2,4} = \sqrt{(0,656 - 0,313)^2 + (0,724 - 0,667)^2 + (0,864 - 0,933)^2} = 0,152$$

2. Hitung nilai SSW (*Sum of square within*)

$$SSW_1(C_1) = \frac{0,193 + 0,119 + 0,229 + 0,151}{4} = 0,173$$

$$SSW_2(C_2) = \frac{0,103 + 0,086 + 0,152 + 0,066 + 0,245 + 0,125 + 0,161 + 0,230}{8} = 0,146$$

$$SSW_3(C_3) = \frac{0,158 + 0,056 + 0,197 + 0,292 + 0,228 + 0,146 + 0,063 + 0,086}{8} = 0,153$$

3. Mencari nilai *Sum of square between cluster* (SSB)

$$SSB_{1,2} = C_{1,2} = \sqrt{(0,327 - 0,656)^2 + (0,810 - 0,724)^2 + (0,850 - 0,864)^2} = 0,34$$

$$SSB_{1,3} = C_{1,3} = \sqrt{(0,327 - 0,286)^2 + (0,810 - 0,330)^2 + (0,850 - 0,815)^2} = 0,483$$

$$SSB_{2,3} = C_{2,3} = \sqrt{(0,656 - 0,286)^2 + (0,724 - 0,330)^2 + (0,864 - 0,815)^2} = 0,543$$

4. Hitung nilai Rasio

$$R_{1,2} = \frac{0,173 + 0,146 + 0,153}{0,34} = 0,938$$

$$R_{1,3} = \frac{0,173 + 0,146 + 0,153}{0,483} = 0,675$$

$$R_{2,3} = \frac{0,173 + 0,146 + 0,153}{0,543} = 0,551$$

5. Menentukan nilai DBI

Untuk mencari nilai DBI yaitu hasil dari Rmax dijumlahkan dan dibagi dengan jumlah data. Maka menghasilkan nilai dari DBI dapat dilihat pada Tabel 4.

Tabel 4. Hasil akhir nilai DBI

No	1	2	3	R max	DBI
1	0	0.938	0.675	0.938	
2	0.938	0	0.551	0.938	0,85
3	0.675	0.551	0	0.675	

4. KESIMPULAN

Berdasarkan hasil penelitian yang dilakukan dapat disimpulkan bahwa proses pengelompokan (*clustering*) penggunaan internet pada siswa persekolah dasar menggunakan algoritma *K-Means*, yaitu tahap pertama yang dilakukan dengan menentukan jumlah *cluster*. Pada penelitian ini, *cluster* yang ditentukan ada 3, yaitu *cluster* pertama (C1) dikategorikan sebagai kurang aktif *cluster* ke 2 dikategorikan sebagai cukup aktif dan *cluster* ke 3 dikategorikan sebagai aktif. Selanjutnya menentukan titik pusat *cluster* (*centroid* awal) disesuaikan dengan jumlah variabel. Jika jumlah variabel dan *centroid* awal sudah ditentukan maka lanjut ke perhitungan menggunakan rumus pengukuran jarak *Euclidean Distance*. Setelah menghitung dengan rumus pengukuran jarak, hasilnya di kelompokkan berdasarkan nilai jarak yang diminimumkan hingga objek pada *cluster* tidak berpindah maka hitungan menggunakan algoritma *k-means* dianggap selesai. Dengan memperoleh hasil akhir yaitu Sekolah yang masuk pada *cluster* satu (C1) atau sekolah yang dikategorikan kurang aktif berinternet ada 4 Sekolah

yaitu SD Negeri 11, SD Negeri 8, SD Negeri 9 dan SD Pertiwi 1. Sekolah yang masuk pada *cluster* dua (C2) atau sekolah yang dikategorikan cukup aktif berinternet ada 8 yaitu SDN 1, SDN 10, SDN 12, SDN 15, SDN16, SDN 17, SDN 19 dan SDN 4 dan Sekolah yang masuk pada *cluster* tiga (C3) atau sekolah yang dikategorikan aktif berinternet ada 8 yaitu SDN 13, SDN 14, SDN 18, SDN 2, SDN 3, SDN 5, SDN 6 dan SDN 7.

DAFTAR PUSTAKA

- APJII (Asosiasi Penyelenggara Jasa Internet Indonesia). (2016) "Data Statistik Pengguna Internet Indonesia"
- ASTRID KURNIA SHERLYANITA, NUR AINI RAKHMAWATI, (2016) "Pengaruh dan Pola Aktivitas penggunaan internet serta media sosial pada siswa SMPN 52 Surabaya" e-journal.unair.ac.id.
- DESI PIBRIANA, DESY IBA RICOIDAPADA, (2017) "Analisis Pengaruh Internet Terhadap minat belajar mahasiswa" journalnasional.ump.ac.id.
- DINAS PENDIDIKAN DAN KEBUDAYAAN KOTA TERNATE. (2017) "Data Sekolah Dasar Perkecamatan Di Kota Ternate" Disdiknas.ternatekota.go.id
- EMIR MAULUDI HUSNI, AGUS FATULLOH,(2016) "Kategorisasi pengguna internet di kalangan pelajar SD dan SMP Menggunakan metode Twostep Cluster" journal.uui.ac.id.
- KAMARGA, H. (2002). Belajar sejarah melalui e-learning: Alternatif Mengakses Sumber Informasi Kesejarahan. Jakarta: Inti Media.
- LANI SIDHARTA (1996) "Penggunaan Internet dan informasi internet" elib.unikom.ac.id.
- PUSPITA ADIYANI CHANDRA. (2013) "Penggunaan internet Pada Anak-Anak Sekolah Usia 6-12 Tahun Di Surabaya" journal unair.ac.id.
- PURNAMASARI. (2013) "Machine Learning cabang dari Kecerdasan Buatan" lintang staff. Gunadarma.ac.id
- SYARIFUDDIN N KAPITA, DKK (2022), The Correlation Between Intensity and Objectives of Using The Internet Towards Students Learning Achievement Using Apriori Algorithm, *MATEC Web of Conferences*.