

Klastering Spesifikasi Smartphone Menggunakan Algoritma Self Organizing Map

Isnain Nabawi*¹, Kusri², Sudarmawan³

^{1,2,3}Program Studi S2 Teknik Informatika, Program Pascasarjana, Universitas Amikom Yogyakarta
e-mail: ¹isnain.1299@amikom.com, ²kusri@amikom.ac.id, ³sudarmawan@amikom.ac.id

Abstrak (Bahasa Indonesia)

Smartphone merupakan alat komunikasi modern yang saat ini sudah banyak dimiliki oleh masyarakat Indonesia. Tidak ada pengelompokan yang pasti dari banyaknya smartphone yang tersebar saat ini sehingga konsumen cukup kesulitan memilih smartphone dan produsen pun perlu mengelompokkan sendiri pasarnya. Pengelompokan smartphone dirasa perlu untuk membantu konsumen untuk mengetahui kelas smartphone yang akan dibelinya berdasarkan spesifikasi yang diinginkan. Pengelompokan pada penelitian ini menggunakan metode Self-Organizing Maps yang dieksekusi menggunakan Bahasa pemrograman Python. Hasil pengelompokan dievaluasi menggunakan 3 metode evaluasi yaitu Silhouette Score, Davies Bouldin Index dan Calinski-Harabasz Index. Hasil menunjukkan bahwa hasil pengelompokan terbaik yaitu pengelompokan yang menghasilkan 2 kelas dengan perolehan Silhouette Score sebesar 0.408, DBI sebesar 1.045, dan CHI sebesar 449.826.

Kata kunci—Spesifikasi Smartphone, Self-organizing Maps, Silhouette Score, Davies Bouldin Index, Calinski-Harabasz Index.

Abstrack (Bahasa Inggris)

Smartphones are a modern form of communication that are now widely owned by the people of Indonesia. There is no clear classification of the many smartphones that are currently available, making it difficult for consumers to choose a smartphone and manufacturers must classify their own markets. Classification of smartphones is necessary to help consumers to know the class of smartphone they will buy based on the desired specifications. The classification in this research used the Self-Organizing Maps method, which was executed using the Python programming language. The classification results were evaluated using 3 evaluation methods, namely the Silhouette Score, Davies Bouldin Index, and Calinski-Harabasz Index. The results show that the best classification is the classification that results in 2 classes with a Silhouette Score of 0.408, a DBI of 1.045, and a CHI of 449.826.

Keywords— Smartphone Specification, Self-organizing Maps, Silhouette Score, Davies Bouldin Index, Calinski-Harabasz Index.

1. PENDAHULUAN

Saat ini smartphone merupakan alat komunikasi yang hampir seluruh lapisan masyarakat memilikinya. Melalui viva.co.id Menteri Komunikasi dan Informatika (Menkominfo) Johnny G Plate mengatakan bahwa jumlah pemilik smartphone atau ponsel pintar di Indonesia mencapai 345,3 juta, lebih banyak dari total penduduk yang mencapai 271,35 juta jiwa [1]. Ini artinya, hampir seluruh lapisan masyarakat memiliki smartphone, dari lapisan atas bahkan hingga lapisan bawah. Kebutuhan akan smartphone selain sebagai media komunikasi, saat ini juga semakin banyak layanan masyarakat juga dapat menggunakan smartphone mulai dari pendidikan, surat-menyurat, bahkan perbankan.

Pertumbuhan pengguna smartphone ini menyebabkan semakin banyaknya produk smartphone yang beredar di pasaran dengan berbagai macam spesifikasi yang ditawarkan [2]. Penelitian ini bertujuan untuk mengelompokkan berbagai macam spesifikasi smartphone ke dalam beberapa kategori atau kluster, sehingga dapat memudahkan para pengguna smartphone dalam memilih produk smartphone yang sesuai dengan kebutuhan dan budget yang dimilikinya. Klastering dapat dilakukan dengan menggunakan metode-metode seperti K-Means atau Hierarchical Clustering untuk mengelompokkan data spesifikasi smartphone berdasarkan beberapa kriteria seperti harga, kapasitas memori, kamera, dan sebagainya. Dengan demikian, diharapkan dapat membantu para pengguna smartphone dalam memilih produk yang sesuai

dengan kebutuhan dan budget yang dimilikinya, serta menjadi bahan pertimbangan bagi produsen smartphone dalam menentukan strategi pemasaran produknya.

Tidak ada klasifikasi baku untuk smartphone yang ada saat ini. Beberapa media memiliki istilah masing-masing untuk mengklasifikasi smartphone. Misalkan saja website kontan.co.id mengelompokkan smartphone menjadi 4 kelompok yaitu: entry level; mid-range; high-end; dan flagship [3]. Sedangkan kompas.com hanya mengelompokkan smartphone menjadi 3 kelompok yaitu: entry level, middle-level, dan Flagship. Maka dari itu perlu adanya penelitian untuk dapat mengelompokkan smartphone berdasarkan spesifikasi yang dimilikinya dan mengembangkannya untuk dapat memberi saran dalam memilih smartphone. Bahkan brand juga bisa menentukan smartphone tertentu akan masuk di kelompok mana.

Penelitian lain dalam mengenai processing data spesifikasi smartphone adalah penelitian oleh Rahman, N.T., pada tahun 2020. Penelitian tersebut menggunakan Simple Multi Attribute Rating dalam metode klasifikasinya. Penelitian tersebut hanya menggunakan 1 pengguna saja dan menghasilkan rekomendasi 1 alternatif pilihan saja [4]. Penelitian terkait klustering juga dilakukan oleh Mawaddah Harahap, dkk. pada tahun 2021. Penelitian yang dilakukan oleh mawaddah tersebut bertujuan untuk mengetahui performa metode LVQ dalam memprediksi hasil produksi kelapa sawit. Hasil prediksi menyatakan bahwa produksi terbaik diperoleh pada Epoch 5000 pada tahun 2010 dan 2011 dengan presentasi akurasi sebesar 58,3%. LVQ [5]. Pada penelitian tersebut data langsung digunakan untuk melakukan klasifikasi. Data sudah dikelompokkan dari data aslinya. Data juga tidak dilakukan normalisasi terlebih dahulu. Sedangkan penelitian yang akan dilakukan yaitu menggunakan SOM untuk pengelompokan (clustering).

Penelitian mengenai aplikasi Jaringan Syaraf Tiruan Kohonen Self Organizing Maps juga dilakukan oleh Fitriati pada tahun 2020 untuk memprediksi Kualitas Air. Penelitian ini mengelompokkan seluruh titik pantau kualitas air Kali Surabaya dengan jumlah cluster yang terbentuk dimulai dari 2 hingga 4 cluster berdasarkan kategori pembagian status mutu air. Dari pengelompokan titik pantau ini kemudian dilakukan penetapan status mutu kualitas air selama 14 periode. Uji coba program dapat diperoleh kesimpulan bahwa algoritma Kohonen-SOM dan LVQ dapat mengenali pola dan mampu mencocokkan anggota kelompok titik pantau dengan parameter learning rate minimal 0,000001 diperoleh nilai mean square error pada jaringan LVQ lebih kecil dibandingkan dengan jaringan Kohonen-SOM [6]. Perlunya pengujian lebih luas sesuai dengan periode yang digunakan supaya lebih detail. Kelemahannya perhitungan nilainya sangat kecil sehingga kevalidan mutu kualitas airnya diragukan.

Peneliti memilih algoritma Self Organizing Maps untuk mengelompokkan smartphone berdasarkan spesifikasinya. Hasil dari pengelompokan nantinya akan dievaluasi menggunakan beberapa metode evaluasi. Hasil penelitian ini nanti diharapkan dapat menjadi pedoman peneliti lainnya dalam mengembangkan sistem clustering maupun sebagai dasar dalam pengklasifikasian smartphone yang sudah tersebar di pasaran.

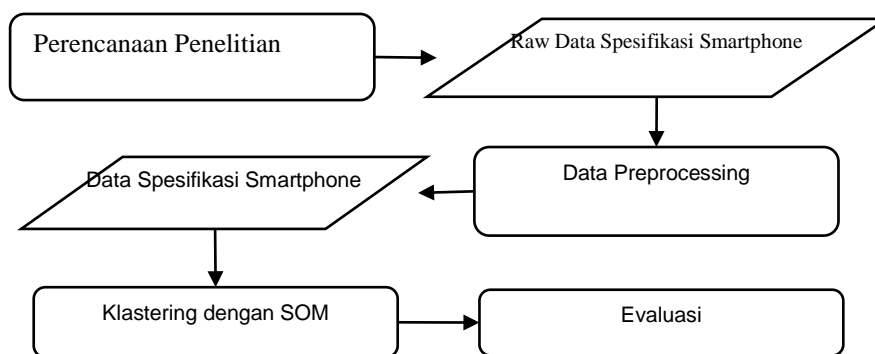
2. METODE PENELITIAN

Penelitian ini diawali dengan perencanaan penelitian. Perencanaan penelitian adalah tahapan untuk mempelajari masalah. Masalah-masalah ini sudah tertuang dalam pendahuluan. Perencanaan penelitian ini juga termasuk melakukan studi tentang penelitian relevan lain yang sudah pernah dilakukan. Sehingga penelitian ini bisa menyempurnakan penelitian-penelitian tersebut. Berdasarkan masalah dan studi penelitian lain didapatkan sebuah ide penelitian tentang klustering data spesifikasi smartphone. Mengumpulkan data spesifikasi smartphone juga termasuk dalam perencanaan penelitian. Data yang digunakan adalah data spesifikasi smartphone dari website GSMarena yang telah diunduh dan dibagikan melalui platform Kaggle hingga tahun 2020. Data terdiri dari 116 merk smartphone dan lebih dari 10.000 model dengan 86 fields spesifikasi. Data bersifat public dan diupload oleh username "Mohit Sainani".

Data smartphone yang digunakan adalah campuran antara numerik dan non numerik. Beberapa kolom juga memiliki beberapa gabungan spesifikasi sehingga harus dilakukan

preprocessing data terlebih dahulu. Dalam tahap ini ada 2 proses utama yaitu proses filtering dan normalisasi data. Proses filter data ini adalah proses untuk menghilangkan data yang rancu atau terdapat missing value. Dalam penelitian ini data yang rancu atau missing value akan dihilangkan. Proses normalisasi data pada penelitian ini menggunakan metode min-max scale adalah dengan mentransformasikan data kedalam range 0-1 [7]. Tahapan ini memiliki output berupa data yang sudah siap olah. Data yang sudah siap olah dari tahapan sebelumnya kemudian dibuatlah program yang memuat tahapan pelatihan dan pengujian Algoritma Self-organizing maps.

Pada tahun 1990 Teuvo Kohonen memperkenalkan algoritma Self-Organizing Maps(SOM). Algoritma ini merupakan algoritma unsupervised learning atau algoritma yang tidak perlu memerlukan pengawasan atau data latih [8]. Peneliti menggunakan Python dalam pengolahan data ini. Library SOM pada Python memiliki beberapa parameter. Untuk mendapatkan hasil yang terbaik beberapa variasi nilai parameter akan dilakukan dan saling dibandingkan. Hasil dari beberapa variasi nilai parameter tersebut akan dievaluasi menggunakan beberapa metode evaluasi kluster. Algoritma SOM akan dievaluasi menggunakan metode nilai Silhouette Score, Davis Bouldin Index (DBI), dan Calinski-Harabasz Index (CHI). Hasil evaluasi nantinya akan dilakukan penarikan kesimpulan untuk mendapatkan output penelitian. Untuk lebih lengkapnya alur penelitian dapat dilihat pada gambar Gambar 1.



Gambar 1. Alur Penelitian

Silhouette Score adalah evaluasi kluster yang menilai keketatan dan pemisahan tiap kluster. Metode ini menunjukkan objek mana yang terletak dengan baik di dalam klusternya, dan mana yang hanya berada di suatu tempat di antara kluster [9]. Silhouette Score dihitung menggunakan persamaan 1.

$$S = (b-a) / (\max(a,b)) \quad (1)$$

Silhouette Score adalah "S". Jarak rata-rata antara sampel dan semua titik lain di kelas yang sama adalah "a". Sedangkan b adalah jarak rata-rata antara sampel dan semua titik lain di cluster terdekat berikutnya.

Davies-Bouldin Index adalah ukuran compactness dan pemisahan cluster dalam suatu dataset. Indeks ini didefinisikan sebagai rata-rata kemiripan antar cluster, dimana kemiripan antara dua cluster didefinisikan sebagai nilai maksimum dari rasio jarak dalam cluster terhadap jarak antar cluster untuk setiap titik di satu cluster ke titik terdekatnya di cluster lain. Dengan kata lain, indeks Davies-Bouldin mengukur compactness cluster dan pemisahan antar cluster. Indeks Davies-Bouldin yang lebih rendah menunjukkan bahwa cluster lebih kompak dan lebih terpisah satu sama lain, yang biasanya dianggap sebagai sifat yang baik untuk sebuah algoritma clustering [10].

Calinski-Harabasz Index adalah salah satu metrik yang digunakan untuk mengevaluasi kualitas suatu algoritma clustering. Indeks ini dihitung dengan menggunakan rasio antara variasi inter-cluster dengan variasi intra-cluster. Varian intra-cluster mengukur variasi di dalam setiap luster, sedangkan varian inter-cluster mengukur variasi antar cluster. Indeks Calinski-Harabasz yang lebih tinggi menunjukkan bahwa cluster lebih terpisah satu sama lain dan lebih kompak,

yang dianggap sebagai sifat yang baik untuk sebuah algoritma clustering. Sehingga semakin besar nilai CH semakin baik atau optimum jumlah klasternya [11]. Indeks ini sering digunakan untuk mengevaluasi kualitas algoritma clustering dalam konteks unsupervised learning.

3. HASIL DAN PEMBAHASAN

Penelitian ini melakukan clustering pada data spesifikasi smartphone. Clustering merupakan suatu proses pengelompokan data dari hasil perekaman atau pengamatan yang memiliki kriteria-kriteria yang sama atau memiliki kemiripan antar objek [12]. Data spesifikasi smartphone didapatkan dari sumber data terbuka. Sumber data terbuka tersebut adalah dari website GSMarena yang telah diunduh dan dibagikan melalui platform Kaggle hingga tahun 2020. Data bersifat public dan diupload oleh username “Mohit Sainani”. Data yang diunduh tersebut kemudian disebut dengan Raw Data. Raw Data tidak bisa langsung diproses oleh perhitungan matematika karena masih terdiri dari campuran angka, huruf, dan data kosong. Sehingga sebelum dilakukan proses pengolahan data, proses yang dilakukan selanjutnya adalah melakukan Data Filtering.

Raw Data spesifikasi smartphone yang diunduh terdiri dari 10680 baris termasuk dengan header dan 86 kolom dengan 2 kolom berisi identitas smartphone. Kolom/fields yang tersedia dalam Raw Data tersebut dapat dilihat pada Tabel 1. Raw data selanjutnya akan dilakukan filtering terlebih dahulu dengan cara:

- 1) Menyeleksi nilai-nilai yang kosong/tidak berarti. Seleksi nilai-nilai yang kosong ini dilakukan dengan bantuan Google Spreadsheet melalui fitur filter. Nilai yang kosong/tidak berarti ini seperti: “no”, “-“, (spasi), (null), dll.
- 2) Memilih fields yang akan digunakan. Karena cukup banyak fields yang diberikan data, maka penulis menyeleksi field yang digunakan adalah bertipe data numerik dan berdasarkan pilihan pribadi dari peneliti.

Setelah menyeleksi nilai-nilai yang kosong, data yang tersisa adalah 718 baris. Fields yang dipilih dapat dilihat pada Tabel 2

Tabel 1. Raw Data Fields Spesifikasi Smartphone.

No	Fields
1	oem
2	model
3	network_technology
...	...
84	network
85	battery_talk_time
86	battery_stand.by

Tabel 2. Fields Spesifikasi Smartphone yang digunakan dalam penelitian

No	Fields
1	body_weight
2	display_size
3	battery_capacity
4	total_cpu_speed
5	internal_memory
6	ram
7	main_camera_resolution

Setelah dilakukan pemilihan fields yang digunakan, langkah selanjutnya adalah melakukan normalisasi data. Normalisasi data ini dilakukan dengan metode min-max scale. Cara kerja min-max scale adalah dengan nilai sebuah fitur dikurangi nilai minimum dari fitur tersebut kemudian dibagi dari rentang nilai keseluruhan [13]. Data hasil normalisasi menggunakan metode ini memiliki rentang 0-1.

Peneliti menggunakan Bahasa pemrograman python dalam melakukan pemrosesan data menggunakan algoritma Self-organizing Maps. Peneliti menggunakan library sklearn_som. Library tersebut dapat didapatkan pada tautan <https://pypi.org/project/sklearn-som/>. Pada library tersebut terdapat parameter inisiasi sebelum menjalankan algoritma SOM. Parameter-parameter tersebut dapat dilihat pada Tabel 3.

Dari ke-7 parameter pada Tabel 3 yang peneliti ubah hanya parameter m,n, dan dim karena parameter tersebut menyesuaikan keadaan data. Perkalian antara parameter m dan n akan menghasilkan ruang yang akan ditempati oleh neuron pemenang atau jumlah kluster. Jumlah kluster akan mengikuti ukuran ruang yang terbentuk [14]. Dataset yang sudah dilakukan preprocessing menghasilkan data sebanyak 718 baris. Jumlah kluster maksimal adalah 1 anggota untuk setiap kluster yaitu 718 kluster. Untuk menentukan perkalian lattice yaitu dengan cara mencari akar dari kluster maksimal kemudian dibulatkan keatas. Akar dari 718 adalah 26,79 maka perkalian lattice adalah 27x27 dengan ruang terbentuk maksimal adalah 729. Variasi perkalian lattice untuk membentuk ruang dan jumlah kluster yang terbentuk dapat dilihat pada Tabel 4. Dilihat dari tabel tersebut gap terbesar adalah pada dimensi 22x27.

Tabel 3. Parameter Inisialisasi SOM

No	Parameter	Type Data	Nilai Default	Deskripsi
1	m	integer	3	Dimensi vertikal SOM (tinggi lattice)
2	n	integer	3	Dimensi horizontal SOM (lebar lattice)
3	dim	integer	3	Jumlah fitur yang digunakan
4	lr	float	1	Jumlah step dalam pembaharuan bobot SOM
5	sigma	float	1	Parameter ini digunakan untuk menentukan besarnya perubahan untuk setiap bobot. Semakin tinggi sigma semakin besar perubahan di setiap bobot.
6	max_iter	int	3000	Parameter ini digunakan untuk memberi batasan iterasi data training.
7	random_state	int	None (0)	Parameter ini digunakan untuk memberi bobot SOM awal.

Tabel 4. Jumlah kluster terbentuk berdasarkan ukuran dimensi

Dimensi	Ukuran Ruang	Kluster Terbentuk	Gap
1x27	27	13	14
2x27	54	54	0
3x27	81	81	0
4x27	108	94	14
5x27	135	129	6
6x27	162	152	10
7x27	189	189	0
8x27	216	208	8
9x27	243	205	38
10x27	270	222	48
11x27	297	290	7
12x27	324	318	6
13x27	351	351	0
14x27	378	286	92
15x27	405	306	99
16x27	432	374	58
17x27	459	370	89
18x27	486	418	68
19x27	513	497	16
20x27	540	479	61
21x27	567	556	11
22x27	594	394	200
23x27	621	481	140
24x27	648	648	0
25x27	675	593	82

Dimensi	Ukuran Ruang	Klaster Terbentuk	Gap
26x27	702	690	12
27x27	729	486	243

Setelah mendapatkan nilai variasi klaster berdasarkan ruang yang terbentuk, hal selanjutnya adalah melakukan evaluasi terhadap hasil tiap jumlah klaster. Pada evaluasi ini peneliti mencoba menggunakan 3 metode evaluasi yaitu Silhouette Score, Davis Bouldin Index (DBI), dan Calinski-Harabasz Index (CHI). Untuk melakukan perhitungan ke-3 metode evaluasi tersebut, peneliti menggunakan library Sklearn.

Pada library sklearn untuk menggunakan evaluasi silhouette_score berikut adalah baris program yang dibuat:

```
# Mengimpor fungsi silhouette_score dari library sklearn
from sklearn.metrics import silhouette_score
# Inialisasi fungsi silhouette score
score = silhouette_score(gsm_spec, predictions, metric='euclidean')
# memasukkan hasil perhitungan ke variabel report untuk ditampilkan.
report = report + str('\t%.3f' % score)
```

Nilai Silhouette score yang dihasilkan akan berkisar antara -1 sampai 1. Semakin mendekati 1, berarti semakin baik kinerja dari algoritma klaster tersebut dalam mengelompokkan data [15]. Sebaliknya, semakin mendekati -1 berarti semakin buruk kinerja dari algoritma klaster tersebut.

Library sklearn untuk menggunakan Calinski-Harabasz Index (CHI) dapat dilihat pada baris program berikut:

```
# Mengimpor fungsi calinski_harabasz_score dari library sklearn
from sklearn import metrics
# Inialisasi fungsi silhouette score
score = metrics.calinski_harabasz_score(gsm_spec, predictions)
# memasukkan hasil perhitungan ke variabel report untuk ditampilkan.
report = report + str('\t%.3f' % score)
```

Nilai CHI yang dihasilkan akan semakin besar semakin baik kinerja dari algoritma klaster tersebut dalam mengelompokkan data. Nilai ini bergantung pada jumlah klaster yang digunakan, sehingga untuk mengevaluasi hasil klaster yang berbeda, kita perlu menghitung CH Index untuk masing-masing hasil klaster tersebut dan membandingkannya.

Davis Bouldin Index (DBI) juga terdapat pada library sklearn sehingga untuk menggunakannya adalah dengan baris program berikut:

```
# Mengimpor fungsi davis_bouldin_score dari library sklearn
from sklearn.metrics import davies_bouldin_score
# Inialisasi fungsi davies_bouldin_score
score = davies_bouldin_score(gsm_spec, predictions)
# memasukkan hasil perhitungan ke variabel report untuk ditampilkan.
report = report + str('\t%.3f' % score)
```

Indeks ini mengukur seberapa baik pengelompokan yang terbentuk, dengan memperhitungkan jarak antara cluster dan variabilitas di dalam cluster. Berbeda dengan CHI, nilai DBI yang semakin mendekati 0 menunjukkan jumlah klaster yang paling optimum [16]. Hasil evaluasi klaster dengan 3 metode evaluasi tersebut dapat dilihat pada Tabel 5.

Berdasarkan hasil para Tabel 5 tersebut, jumlah klaster terbaik menurut Silhouette Score dan Cilinski-Harabasz Index adalah jumlah klaster 27 yang terbentuk dari perkalian lattice 1x27 dengan skor 0.127 untuk Silhouette Score dan 242997 untuk Cilinski-Harabasz Index. Sedangkan menurut Davis Bouldin Score, klaster terbaik adalah jumlah klaster 81 yang terbentuk dari perkalian lattice 3x27 dengan skor 1665.

Tabel 5. Evaluasi hasil kluster pada rentang ukuran ruang 27-729

Dimensi	Ukuran Ruang	Kluster terbentuk	Silhouette Score	Davis Bouldin Score	Calinski-Harabasz Index
1x27	27	13	0.172	1.58	242.997
2x27	54	54	0.154	1.595	208.383
3x27	81	81	0.14	1.665	161.693
4x27	108	94	0.135	1.47	146.428
5x27	135	129	0.163	1.468	145.032
6x27	162	152	0.146	1.404	118.758
7x27	189	189	0.155	1.531	125.308
8x27	216	208	0.143	1.422	130.468
9x27	243	205	0.173	1.362	154.564
10x27	270	222	0.137	1.512	132.423
11x27	297	290	0.11	1.526	114.814
12x27	324	318	0.127	1.518	125.259
13x27	351	351	0.117	1.412	137.173
14x27	378	286	0.15	1.359	131.615
15x27	405	306	0.139	1.44	136.46
16x27	432	374	0.102	1.532	130.948
17x27	459	370	0.132	1.453	118.028
18x27	486	418	0.164	1.359	126.258
19x27	513	497	0.127	1.393	145.132
20x27	540	479	0.15	1.359	108.965
21x27	567	556	0.168	1.318	140.706
22x27	594	394	0.129	1.381	131.201
23x27	621	481	0.155	1.386	134.078
24x27	648	648	0.127	1.501	135.684
25x27	675	593	0.156	1.404	137.953
26x27	702	690	0.142	1.457	137.049
27x27	729	486	0.133	1.377	117.35

Hasil dari Tabel 5 menunjukkan bahwa hasil terbaik masih berada di rentang 1x27 kluster atau di jumlah kluster terendah. Karena hasil terbaik berada di kluster terendah maka perlu mencari jumlah kluster optimum dalam rentang maksimal 27 yaitu dengan mencari akar 27. Akar 27 adalah 5,19. Peneliti akan mencari jumlah kluster terbaik dengan ukuran maksimal 5x6. Hasil pencarian tersebut dapat dilihat pada Tabel 6.

Tabel 6 menunjukkan bahwa Silhouette Score terbaik adalah dengan nilai 0.408 pada jumlah kluster 2, DBI terbaik dengan nilai 1.045 pada jumlah kluster 2, dan CHI terbaik dengan nilai 449.826 pada jumlah kluster 2. Semua metode evaluasi yang digunakan memberikan output terbaik yang sama yaitu pada jumlah kluster 2.

Tabel 6. Evaluasi hasil kluster pada rentang ukuran ruang 2-27

Dimensi	Ukuran ruang	Kluster terbentuk	Silhouette Score	Davis Bouldin Index	Calinski-Harabasz Index
2x1	2	2	0.408	1.045	449.826
3x1	3	3	0.31	1.494	325.294
4x1	4	4	0.319	1.176	418.464
5x1	5	5	0.306	1.218	413.329
6x1	6	6	0.287	1.341	348.001
2x2	4	4	0.364	1.206	335.639
3x2	6	6	0.221	1.373	314.012
4x2	8	8	0.168	1.541	249.099
5x2	10	10	0.223	1.482	224.909
6x2	12	12	0.226	1.274	263.992
2x3	6	6	0.27	1.538	308.433
3x3	9	9	0.227	1.477	227.476
4x3	12	12	0.165	1.553	240.539
5x3	15	15	0.123	1.513	149.367
6x3	18	18	0.158	1.516	201.679
2x4	8	8	0.169	1.529	270.903
3x4	12	12	0.205	1.542	236.233
4x4	16	16	0.152	1.463	172.937
5x4	20	20	0.142	1.588	185.786
6x4	24	24	0.141	1.491	197.501
2x5	10	10	0.155	1.414	250.863
3x5	15	15	0.129	1.679	212.757
4x5	20	20	0.149	1.428	195.01
5x5	25	25	0.158	1.625	178.438
6x5	30	30	0.158	1.505	172.711

4. KESIMPULAN

Smartphone merupakan perangkat yang hampir seluruh masyarakat Indonesia menggunakannya. Tidak ada pengelompokan yang pasti untuk smartphone-smartphone tersebut. Peneliti menggunakan metode pengelompokan Self-organizing maps untuk mengelompokkan data smartphone. Field yang digunakan dalam pengelompokan adalah `body_weight`, `display_size`, `battery_capacity`, `total_cpu_speed`, `internal_memory`, `ram`, dan `main_camera_resolution`. Peneliti menggunakan data Hasil pengelompokan tersebut dievaluasi menggunakan 3 metode evaluasi yaitu Silhouette Score, Davies Boildin Index dan Cilinski Harabaz Index. Hasil evaluasi terbaik menunjukkan hasil yang sama yaitu pada 2 jumlah kluster. Silhouette Score terbaik adalah dengan nilai 0.408, DBI terbaik dengan nilai 1.045, dan CHI terbaik dengan nilai 449.826.

Hasil evaluasi kluster menunjukkan hasil terbaik adalah 2 jumlah kluster sedangkan pasar smartphone cukup luas. Jumlah pengelompokan sebaiknya tidak hanya mempertimbangkan dari data smartphone saja tapi juga harus ada riset pasar yang lebih

mendalam. Penambahan metode evaluasi kluster bisa juga menjadi solusi untuk menguatkan pemilihan jumlah kluster yang terbaik.

DAFTAR PUSTAKA

- [1] U. Lazuardhi, "Jumlah Ponsel di Indonesia Lebih Banyak dari Populasi Penduduk," *PT Visi Media Asia Tbk*, 12 Agustus 2021. [Online]. Available: <https://www.viva.co.id/digital/digilife/1395151-jumlah-ponsel-di-indonesia-lebih-banyak-dari-populasi-penduduk>. [Diakses 12 12 2022].
- [2] Irfan, Aswar and Erviana, "HUBUNGAN SMARTPHONE DENGAN KUALITAS TIDUR REMAJA DI SMA NEGERI 2 MAJENE," *Journal of Islamic Nursing*, vol. 5, no. 2, pp. 95-100, 2020.
- [3] R. P. Gozal, "Ini Tingkatan Smartphone Berdasarkan Harga & Spesifikasi," *PT Grahanusa Mediatama*, 15 April 2021. [Online]. Available: <https://seremonia.kontan.co.id/news/ini-tingkatan-smartphone-berdasarkan-harga-spesifikasi>. [Diakses 24 12 2022].
- [4] N. T. Rahman and I. N. Kholifah, "SISTEM PENDUKUNG KEPUTUSAN UNTUK PEMILIHAN SMARTPHONE DENGAN MENGGUNAKAN METODE SMART (SIMPLE MULTY ATTRIBUTE RATING)," *JURNAL FASILKOM*, vol. 10, no. 3, pp. 184-191, 2020.
- [5] M. Harahap, A. Mutia, D. B. M. Simatupang and B. S. Gurning, "Implementasi Algoritma Learning Vector Quantization (LVQ) Pada Prediksi Produksi Tandan Buah Segar Pada Perkebunan Kelapa Sawit," *Jurnal SAINTIKOM (Jurnal Sains Manajemen Informatika dan Komputer)*, vol. 20, no. 2, pp. 124-129, 2021.
- [6] S. R. Fitriatien, *Aplikasi Jaringan Syaraf Tiruan Kohonen Self Organizing Maps dan Learning Vector Quantization pada Data Kualitas Air Kali Surabaya*, Surabaya: INSTITUT TEKNOLOGI SEPULUH NOPEMBER, 2015.
- [7] I. Permana and F. N. Salisah, "Pengaruh Normalisasi Data Terhadap Performa Hasil Klasifikasi Algoritma Backpropagation," *JIRSE: Indonesian Journal of Informatic Research and Software Engineering*, vol. 2, no. 1, pp. 67-72, 2022.
- [8] R. Umar, A. Fadlil and R. R. A. Zahra, "Self Organizing Maps (SOM) untuk Pengelompokan Jurusan SMK," *Khazanah Informatika*, vol. 4, no. 2, pp. 131-137, 2019.
- [9] J. Rousseeuw and Peter, "Silhouettes: A graphical aid to the interpretation and validation of cluster analysis," *Journal of Computational and Applied Mathematics*, vol. 20, pp. 53-65, 1987.
- [10] A. A. Aida Fastabiqah Khairati, G. Hertonoa and B. Handari, "Kajian Indeks Validitas pada Algoritma K-Means Enhanced dan K-Means MMCA," *PRISMA, Prosiding Seminar Nasional Matematika*, vol. 2, pp. 161-170, 2019.
- [11] Aliyatussyah'ni and D. R. S. Saputro, "ALGORITME PARTITIONING AROUND MEDOID (PAM) DENGAN CALINSKI-HARABASZ INDEX UNTUK CLUSTERING DATA OUTLIER," *Seminar Nasional Matematika, Geometri, Statistika, dan Komputasi*, pp. 23-29, 2022.

-
- [12] Larose, D. T., Larose and C. D., *Data Mining and Predictive Analytics*. Second Edition., Hoboken, New Jersey.: John Wiley & Sons, 2015.
- [13] A. Ambarwari, Q. J. Adrian and Y. Herdiyeni, "Analisis Pengaruh Data Scaling Terhadap Performa Algoritme Machine Learning untuk Identifikasi Tanaman," *JURNAL RESTI*, vol. 4, no. 1, pp. 117-122, 2022.
- [14] A. F. Solikin, *KLASTERING DATA INTERKOMPARASI ANAK TIMBANGAN MENGGUNAKAN ALGORITMA SELF ORGANIZING MAPS*, Yogyakarta: Universitas Amikom Yogyakarta, 2021.
- [15] M. A. Nahdliyah, T. Widiharah and A. Prahutama, "METODE k-MEDOIDS CLUSTERING DENGAN VALIDASI SILHOUETTE INDEX DAN C-INDEX," *JURNAL GAUSSIAN*, vol. 8, no. 2, pp. 161-170, 2019.
- [16] E. Muningsih, I. Maryani and V. R. Handayani, "Penerapan Metode K-Means dan Optimasi Jumlah Cluster dengan Index Davies Bouldin untuk Clustering Propinsi Berdasarkan Potensi Desa," *Evolusi: Jurnal Sains dan Manajemen*, vol. 9, no. 1, pp. 95-100, 2021.