

## PERBANDINGAN EVALUASI METODE DAVIES BOULDIN, ELBOW DAN SILHOUETTE PADA MODEL CLUSTERING DENGAN MENGUNAKAN ALGORITMA K MEANS

Muhammad Sholeh<sup>1</sup>, Khurotul Aeni<sup>2</sup>

Program Studi Informatika, Fakultas Teknologi Informasi dan Bisnis,  
Institut Sains & Teknologi AKPRIND Yogyakarta<sup>1</sup>

Program Studi Informatika, Fakultas Sains & Teknologi, Universitas Peradaban<sup>2</sup>  
muhash@akprind.ac.id<sup>1</sup>, khaeni988@gmail.com<sup>2</sup>

*Submitted February 1, 2023; Revised April 14, 2023; Accepted June 5, 2023*

### Abstrak

Salah satu model *data mining* yang sering digunakan adalah model *clustering*. Model *clustering* digunakan untuk membuat pengelompokan dari suatu *datasheet*. Pengelompokan data dapat digunakan untuk membedakan data-data yang ada dalam suatu *datasheet*. Banyak pengelompokan yang paling baik dapat dilakukan dengan proses evaluasi model *clustering*. Dalam metode *clustering*, proses evaluasi dapat menggunakan berbagai metode evaluasi. Penelitian akan melakukan pengelompokan hasil survey tentang ulasan tentang destinasi wisata yang terdiri dari 10 kategori. Metode penelitian menggunakan metode *data mining* yaitu *Knowledge Discovery in Database* (KDD). Tahapan dalam KDD diantaranya, seleksi data, membersihkan data, transformasi, proses *data mining* dan evaluasi hasil model. Proses pembuatan model *clustering* menggunakan *datasheet* *tripadvisor\_review.csv*. Proses pengelompokan data menggunakan algoritma K-means. Hasil dari pengelompokan akan diuji dengan melakukan perbandingan metode evaluasi. Metode evaluasi ini digunakan untuk memilih jumlah pengelompokan yang terbaik. Pengujian untuk mendapatkan hasil pengelompokan yang terbaik, dilakukan dengan pengujian mulai dari pengelompokan 2 sampai 14. Metode evaluasi menggunakan metode *Davies Bouldin*, *Elbow* dan *Silhouette*. Hasil penelitian menunjukkan banyaknya pengelompokan *datasheet* dengan ketiga metode evaluasi memberikan rekomendasi pengelompokan sebanyak 2 kelompok.

**Kata kunci :** Clustering, evaluasi, Davies Bouldin, Elbow, Silhouette

### Abstract

*One of data mining model that is often used is the clustering model. The clustering model is used to create a grouping of a datasheet. Data clustering can be used to distinguish data in a datasheet. Many of the best groupings can be done with the clustering model evaluation process. In the clustering method, the evaluation process can use various evaluation methods. The research will cluster the results of a survey on reviews of tourist destinations consisting of 10 categories. The method used is a data mining method, namely Knowledge Discovery in Database (KDD). Stages in KDD include data selection, data cleaning, transformation, data mining process and evaluation of model results. The process of creating a clustering model uses a public datasheet, namely tripadvisor\_review.csv. The data clustering process uses the K-means algorithm. The result of the clustering will be tested by comparing evaluation methods. This evaluation method is used to select the best amount of clustering. Testing to get the best clustering results is conducted by testing from clustering 2 to 15. The evaluation uses Davies Bouldin, Elbow and Silhouette methods. The result shows that the number of datasheet groupings with the three evaluation methods provides recommendations for grouping as many as 2 groups.*

**Keywords:** Clustering, evaluation, Davies Bouldin, Elbow, Silhouette

### 1. PENDAHULUAN

Kepuasan pelanggan terhadap layanan menjadi salah satu indikator keberhasilan

dalam pengelolaan suatu usaha atau kegiatan. Salah satu cara untuk mendapatkan kepuasan dari pelanggan

adalah dengan memberikan survey [1]. Hasil survey tentunya dapat digunakan untuk melihat respon pelanggan. Proses olah data hasil survey dapat menggunakan model data mining, terutama *clustering*.

*Data mining* dapat digunakan dalam suatu pengambilan keputusan. Data mining banyak digunakan dalam melakukan prediksi, klasifikasi, pengelompokan dan asosiasi [2], [3]. Pemilihan model data mining harus disesuaikan dengan datasheet yang diolah. Model regresi digunakan untuk melakukan prediksi atau perkiraan dari suatu data, sedangkan model klasifikasi digunakan untuk melakukan prediksi yang berupa data label yang sudah diklasifikasikan. Model *data mining* yang lain dan digunakan dalam pengelompokan suatu datasheet adalah model *clustering*.

Pembuatan model sangat diperlukan dalam proses analisis data. Penelitian *data mining* sudah banyak dilakukan dengan menggunakan berbagai *datasheet*. Penelitian yang menggunakan model regresi diantaranya memprediksi hasil nilai kuesioner [4], prediksi penjualan mobil [5], prediksi kelulusan mahasiswa [6], prediksi pencegahan hipertensi [7], prediksi biaya asuransi kesehatan [8] dan prediksi status pengiriman barang [9].

Penelitian *data mining* model klasifikasi diantaranya melakukan mengklasifikasikan dengan algoritma K-NN [10] dan menggunakan algoritma *Naïve Bayes* [11].

Model *data mining* yang lain adalah model *clustering*. Model *clustering* digunakan untuk membuat pengelompokan suatu *datasheet*. Hasil pengelompokan dapat digunakan untuk membedakan antar kelompok. Beberapa penelitian yang menggunakan *clustering* diantaranya membuat *cluster* konsumsi produk kosmetik [12], *cluster* data kasus Covid [13], *cluster* data pembayaran transaksi [14] dan *cluster* komoditas perikanan [15].

Penelitian Amanda [12], membuat pengelompokan nama-nama produk yang dapat digunakan untuk melakukan analisis produk yang laku di pasaran dan yang tidak laku di pasaran. Hardiani [13], melakukan pengelompokan berdasar wilayah provinsi menjadi beberapa *cluster* berdasar pada penyebaran Covid-19. Pengelompokan ini bermanfaat untuk mengetahui daerah yang jumlah kasus Covid-19 yang banyak dan yang sedikit. Manalu [14], dalam penelitian yang dilakukan menerapkan metode *clustering* dengan menggunakan algoritma K-Means. Penelitian menghasilkan pengelompokan data-data yang digunakan dalam proses transaksi pembayaran. Banyak pengelompokan pada data proses transaksi pembayaran menghasilkan lima buah dan diharapkan menjadi pertimbangan dalam menentukan strategi perusahaan. Penelitian Maulana [15], menghasilkan analisis berupa produksi hasil komoditas perikanan yang berdasar pada data provinsi dikelompokkan menjadi 3 buah kelompok (*cluster*). Hasil pengelompokan diberi kategori terbesar, sedang, dan terkecil. Adanya pengelompokan dapat menjadi cara untuk mengukur potensi provinsi dalam menjalankan strategi program membangkitkan potensi unggulan provinsi dari segi budidaya perikanan.

Proses menentukan banyak pengelompokan yang paling baik dilakukan dengan evaluasi kinerja dari suatu hasil pengelompokan. Metode atau teknik yang dapat digunakan adalah *Davies Bouldin*, *Elbow*, *Silhouette* dan teknik lainnya. Penelitian yang menggunakan teknik evaluasi kinerja dengan *Davies Bouldin* diantaranya dengan *cluster* data penduduk miskin [16], *cluster* potensi desa [17] dan *cluster* karakteristik kanal *You tube* [18]. Penelitian yang menggunakan teknik *Elbow* dalam proses evaluasi kinerja dilakukan Winarta untuk melakukan *cluster* data narkoba [19], *cluster* data untuk mengukur kesiapan dalam ujian [20] dan *cluster*

untuk pengelompokan sentra industri produksi [21].

Winarta [19], menggunakan teknik *Elbow* untuk melakukan evaluasi hasil pengelompokan pengguna narkoba. Hasil pengujian kinerja dengan metode *Elbow* menghasilkan nilai k yang terbaik adalah 3. Pengujian dilakukan dengan nilai k=2 sampai k=5. Hasil SSE untuk nilai k=2 adalah 1257,862.

Penelitian yang menggunakan teknik *Silhouette* dalam proses evaluasi kinerja dilakukan [22] dengan *cluster* obat dan [23] melakukan penelitian untuk perhitungan jarak K-Means.

Panembonan dalam penelitian yang dilakukan membuat pengelompokan untuk jenis obat yang mirip menjadi satu kelompok tertentu. Hasil pengelompokan dilakukan evaluasi untuk mendapatkan hasil pengelompokan yang terbaik. Proses evaluasi menggunakan teknik *silhouette* dan hasil nilai dari *silhouette coefficient* adalah sebesar 0,4854 [22].

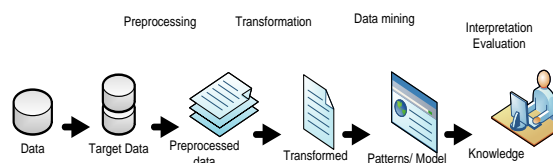
Penelitian lain, melakukan proses evaluasi dengan perbandingan beberapa teknik evaluasi, diantaranya dilakukan Dewi. Dewi dalam penelitiannya melakukan evaluasi dengan membandingkan evaluasi dengan metode *Elbow* dan *Silhouette* [24].

Berdasar pada latar belakang dan tinjauan pustaka di atas, penelitian yang dilakukan akan membuat pengelompokan dengan *datasheet* tripadvisor\_review.csv.

## 2. METODE PENELITIAN

### Proses Data Mining

Metode *data mining* dalam penelitian yang dilakukan menggunakan metode *Knowledge Discovery in Database (KDD)*. Tahapan dalam metode KDD diantaranya melakukan proses seleksi data, membersihkan data, transformasi, proses *data mining* dan evaluasi hasil model. Tahapan *data mining* KDD ditampilkan pada gambar 1.



Gambar 1. Tahapan dalam KDD

### Peralatan Penelitian

Implementasi dari pembuatan model *clustering* menggunakan Python. Python sangat mendukung dalam proses pembuatan data mining [25]. Proses pembuatan model *clustering* memanfaatkan *library* yang sudah tersedia. *Library* yang terkait dengan model clustering yang digunakan diantaranya *sklearn.cluster.KMeans* dan *library* untuk mengevaluasi kinerja.

### Datasheet

Proses pengujian hasil pengelompokan menggunakan *datasheet public* yaitu *tripadvisor\_review.csv* yang diperoleh dari laman

<https://archive.ics.uci.edu/ml/datasets/Travel+Reviews>.

*Datasheet* berisi ulasan perjalanan dari para pengguna. *Datasheet* terdiri dari 980 data dengan 11 *attribute*. Isi data berupa data 0-4 yang merupakan isian yang diisi para wisatawan. Setiap wisatawan mengisi ulasan dengan skor mulai dari 0-4. Skor 4 untuk ulasan sangat baik (4), 3 untuk ulasan baik, 2 untuk ulasan biasa, 1 untuk ulasan buruk dan 0 untuk ulasan sangat jelek. Hasil pengelompokan akan diuji dengan 3 metode atau teknik evaluasi kinerja model *clustering*.

### Model Clustering

Algoritma yang digunakan dalam pembuatan model *clustering* dapat menggunakan algoritma *K-Means*, *K-Medoids*, *DbSCAN* dan lainnya [26]. Model *Clustering* merupakan salah satu model dalam *data mining* yang digunakan untuk melakukan proses pengelompokan dari kumpulan data ke dalam bentuk

pengelompokan yang masing-masing hasil pengelompokan mempunyai kemiripan.

Salah satu algoritma yang dapat digunakan untuk menemukan pengelompokan dalam data adalah dengan menggunakan algoritma *K-Means*.

### Evaluasi Clustering

Proses menentukan banyak kelompok yang ideal dilakukan dengan menggunakan teknik evaluasi model yaitu *Davies-Bouldin index*, metode *Elbow* dan metode *Silhouette*.

### Metode Davies-Bouldin index

*Davies-Bouldin index* (DBI) digunakan untuk melakukan proses evaluasi dari hasil model clustering. DBI melakukan proses pengukuran suatu hasil pengelompokan berdasar pada nilai kohesi dan separasi. Hasil pengelompokan kohesi merupakan jumlah dari kedekatan data terhadap suatu titik pusat (*centroid*) dari suatu hasil pengelompokan dan hasil pengelompokan separasi berdasar pada jarak antar *centroid* dari *cluster* nya.

### Metode Elbow

Metode *Elbow* melakukan proses evaluasi dengan membandingkan hasil antara penjumlahan *cluster* yang akan membentuk suatu titik terakhir dan *cluster* nilai digunakan sebagai model data untuk *cluster* terbaik. Hasil dari proses perhitungan digunakan untuk proses dalam membandingkan jumlah *cluster*.

### Metode Silhouette

Metode *silhouette coefficient* merupakan gabungan dari dua metode yaitu metode *cohesion* yang digunakan untuk mengukur berapa dekat relasi antara satu objek dengan objek lain dalam suatu *cluster*, dan metode *separation* yang berguna untuk mengukur berapa jauh suatu *cluster* yang terpisah dengan *cluster* yang lain

## 3. HASIL DAN PEMBAHASAN

### Data Cleaning

Tahap *data cleaning* dilakukan dengan memeriksa *datasheet* dari data-data yang tidak diperlukan atau membersihkan *datasheet* dari data-data yang tidak diperlukan. Proses yang dilakukan diantaranya melakukan proses pengecekan data kosong, pengecekan daya yang duplikat dan data pencilan [2].

Proses *data cleaning* dilakukan dengan memeriksa *datasheet* dari data yang kosong (*missing value*). Pengecekan data kosong dilakukan dengan memberikan perintah `df.isnull().sum`. Perintah ini akan menjumlahkan data yang mengandung data kosong [3]. Hasil pengecekan *datasheet* yang mengandung *missing value* sudah tidak ada. Hasil pengecekan ditampilkan pada gambar 2.

```
1 #memvalidasi data
2 #cek missing value
3 df.isnull().sum()

User ID      0
Category 1   0
Category 2   0
Category 3   0
Category 4   0
Category 5   0
Category 6   0
Category 7   0
Category 8   0
Category 9   0
Category 10  0
dtype: int64
```

Gambar 2. Proses Pengecekan Data Kosong

Proses pengecekan data yang sama (*duplicate data*) dilakukan dengan memberikan perintah `df.duplicate.sum`. Perintah ini akan menjumlahkan data yang mengandung data yang ganda. Hasil pengecekan data ganda sudah tidak ada. Hasil pengecekan data duplikat ditampilkan pada gambar 3.

```
1 #mengecek data duplikat
2 df.duplicated().sum()
0
```

**Gambar 3. Proses Pengecekan Data Kosong**

Pengecekan data *outlier* (pencilan) dilakukan dengan memeriksa informasi yang dihasilkan dari perintah `df.describe`. Perintah ini akan menampilkan data

statistika dari *datasheet*. Proses pengecekan *datasheet* dari data yang mengandung data pencilan, dilakukan dengan melihat nilai min dan nilai max. Dari nilai min dan max, dapat dilihat apakah ada data yang menyimpang dari nilai yang semestinya ada dalam *datasheet*. Hasil pengecekan dengan perintah *describe*, ditampilkan pada gambar 4.

```
1 #cek apakah ada data yang menyimpang, Lihat nilai max dan min
2 df.describe()
```

	Category 1	Category 2	Category 3	Category 4	Category 5	Category 6	Category 7	Category 8	Category 9	Category 10
count	980.000000	980.000000	980.000000	980.000000	980.000000	980.000000	980.000000	980.000000	980.000000	980.000000
mean	0.893194	1.352612	1.013306	0.532500	0.939735	1.842898	3.180939	2.835061	1.569439	2.799224
std	0.326912	0.478280	0.788607	0.279731	0.437430	0.539538	0.007824	0.137505	0.364629	0.321380
min	0.340000	0.000000	0.130000	0.150000	0.060000	0.140000	3.160000	2.420000	0.740000	2.140000
25%	0.670000	1.080000	0.270000	0.410000	0.640000	1.460000	3.180000	2.740000	1.310000	2.540000
50%	0.830000	1.280000	0.820000	0.500000	0.900000	1.800000	3.180000	2.820000	1.540000	2.780000
75%	1.020000	1.560000	1.572500	0.580000	1.200000	2.200000	3.180000	2.910000	1.760000	3.040000
max	3.220000	3.640000	3.620000	3.440000	3.300000	3.760000	3.210000	3.390000	3.170000	3.660000

**Gambar 4. Proses Pengecekan Data Kosong**

Dari gambar 4 nilai min tidak ada yang di bawah 0 dan maksimal tidak ada yang lebih dari 4, sehingga *datasheet* tidak ada nilai yang menyimpang dari nilai yang normal (0-4).

**Data Selection**

Proses data *selection* digunakan untuk melakukan pemilihan *attribute* atau jumlah data yang digunakan dalam proses *data mining*. Proses data *selection* yang dilakukan diantaranya melakukan pengecekan *type attribute*, menghapus *attribute* yang tidak digunakan dan mengganti nama *attribute*.

Pengecekan *type attribute* digunakan untuk melihat *type data* pada *attribute*. Proses model *clustering*, *type data* yang digunakan harus bertipe bilangan. Hasil proses pengecekan *attribute*, *attribute user id* bertipe *object*, sehingga *attribute user id* tidak digunakan dan *attribute* yang lain sudah bertipe bilangan.

Menghapus *attribute* yang tidak digunakan dapat mempercepat proses *data mining*. Proses menghapus *attribute* menggunakan perintah `df.drop`. Penghapusan *attribute user id* dengan pertimbangan *attribute* ini tidak akan mempengaruhi proses *clustering*

Proses lain yang dilakukan adalah melakukan penggantian nama *attribute*. Penggantian nama *attribute* digunakan untuk memudahkan penamaan dari kolom yang digunakan

**Data Mining**

Tahapan pada proses *data mining* dilakukan dengan membuat model data mining. Model *data mining* yang akan dilakukan adalah dengan model *clustering* dengan menggunakan algoritma K-Means. Tahapan yang dilakukan diantaranya menentukan *library* dan menentukan jumlah *cluster* yang akan dilakukan pengujian. *Library* yang digunakan diantaranya *pandas*, *sklearn* dan *matplotlib*. Algoritma yang digunakan adalah K-means.

Proses *clustering* dilakukan dengan melakukan pengujian dengan jumlah *clustering* mulai dari pengelompokan 2 sampai 14. Hasil pembuatan pengelompokan menghasilkan kelompok-

kelompok yang saling berdekatan. Hasil *cluster* dilakukan evaluasi model dengan teknik tertentu. Contoh hasil pengelompokan pengujian dari K=1 sampai 6, ditampilkan pada tabel 1.

**Tabel 1. Hasil Pengelompokan dari K=2 sampai K=6**

K	Kelompok 0	Kelompok 1	Kelompok 2	Kelompok 3	Kelompok 4	Kelompok 5
2	598	382				
3	234	376	370			
4	234	167	379	209		
5	126	245	234	167	208	
6	167	206	208	28	245	126

**Evaluasi Model**

Evaluasi model *clustering* digunakan untuk mengetahui pengelompokan data yang ideal. Pada penelitian ini, evaluasi *clustering* yang digunakan yaitu *Elbow*, *silhouette* dan *Indeks Davies-Bouldin*. Proses evaluasi digunakan untuk melihat akurasi yang terbaik pada proses *clustering* dan rekomendasi hasil *clustering* yang digunakan.

**Evaluasi Metode Davies Bouldin**

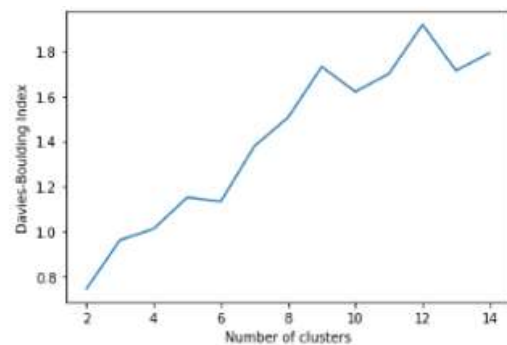
Proses evaluasi dengan metode *Davies Bouldin* dilakukan dengan

membandingkan nilai yang terkecil dari masing-masing pengujian yang dilakukan. Evaluasi model diuji mulai dari k= 2 sampai 14 [17].

Hasil evaluasi dengan menggunakan *Davies Bouldin* dari nilai k= 2 sampai 14, ditampilkan pada gambar 5. Berdasar pada gambar 5, hasil pengelompokan dengan nilai terkecil adalah pada nilai k=2, dengan nilai 0.747, sehingga banyak *cluster* yang direkomendasikan adalah k=2.

```

K = 2 = 0.7473229800702164
K = 3 = 0.9628030655188718
K = 4 = 1.012141755984028
K = 5 = 1.1517992267938912
K = 6 = 1.1332708516259935
K = 7 = 1.3793678135907623
K = 8 = 1.505432956996054
K = 9 = 1.7300390684049356
K = 10 = 1.6198326548184319
K = 11 = 1.69874133974223
K = 12 = 1.916856691319188
K = 13 = 1.7127446222446514
K = 14 = 1.7911386584255289
    
```



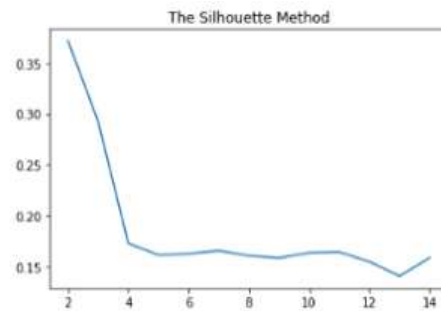
**Gambar 5. Hasil dan Grafik Perhitungan Kinerja dengan Teknik Davies Bouldin**

**Evaluasi Metode Silhouette**

Hasil evaluasi dengan metode *Silhouette* dilakukan dengan membuat beberapa cluster dan membandingkan nilai *Silhouette*. Jumlah *cluster* yang dipilih adalah nilai *Silhouette* yang terbesar. Hasil proses evaluasi dengan metode *Silhouette* ditampilkan pada gambar 6.

Berdasar pada gambar 6, jumlah cluster yang dilakukan pengujian mulai dari k=2 sampai k=14. Hasil pengujian dengan metode *Silhouette* menghasilkan nilai k yang paling besar ada pada nilai k=2 dengan nilai 03718.

K = 2 = 0.3717604495004071  
 K = 3 = 0.29246326429771136  
 K = 4 = 0.17244257965891968  
 K = 5 = 0.16120049326570904  
 K = 6 = 0.16235412238068278  
 K = 7 = 0.1653785768206396  
 K = 8 = 0.16040380860456957  
 K = 9 = 0.15829297213854676  
 K = 10 = 0.16328066094462285  
 K = 11 = 0.16407957888537833  
 K = 12 = 0.1544941987758232  
 K = 13 = 0.1403658740435097  
 K = 14 = 0.15846031855225867



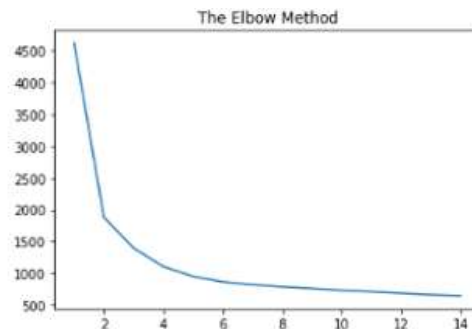
Gambar 6. Hasil dan Grafik Perhitungan Kinerja dengan Teknik *Elbow*

**Evaluasi Metode *Elbow***

Evaluasi dengan menggunakan Metode *Elbow* dilakukan dengan menguji *cluster* mulai dari 2 sampai 14. Jumlah *cluster* dipilih dengan melihat nilai sudut yang

besar dari perbandingan setiap *cluster*. Berdasar pada gambar 8, *cluster* 2 merupakan nilai yang terbaik dari metode *Elbow*.

K = 1 = 4622.8643654081625  
 K = 2 = 1880.4071952833897  
 K = 3 = 1396.2255969344205  
 K = 4 = 1105.3178374664885  
 K = 5 = 951.4395130070807  
 K = 6 = 864.9735599031321  
 K = 7 = 820.9078000250126  
 K = 8 = 787.1786209909179  
 K = 9 = 761.5067853476846  
 K = 10 = 731.912120119049  
 K = 11 = 711.1686278745616  
 K = 12 = 685.5896588825563  
 K = 13 = 662.2828736495667  
 K = 14 = 645.4812901603643



Gambar 8. Hasil dan Grafik Perhitungan Kinerja dengan Teknik *Elbow*

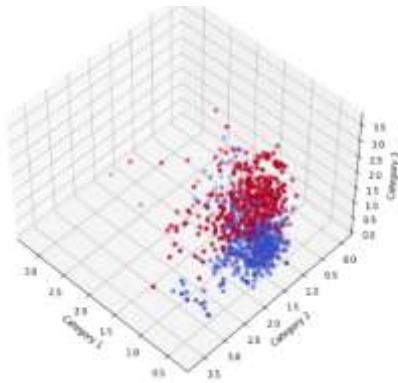
**Hasil pengelompokan**

Sesuai dengan hasil evaluasi dengan 3 metode, banyak pengelompokan yang terbaik adalah 2 pengelompokan. Hasil pengelompokan ini dapat menjadi rekomendasi untuk melakukan pengelompokan dari *datasheet* yang

digunakan. Nama kategori dapat diberi penamaan hasil *cluster* 1 dengan predikat puas dan hasil pada *cluster* 1 dengan predikat tidak puas. Hasil pengelompokan dalam bentuk tabel ditampilkan pada gambar 9 dan pengelompokan dalam bentuk visual ditampilkan pada gambar 10.

	Category 1	Category 2	Category 3	Category 4	Category 5	Category 6	Category 7	Category 8	Category 9	Category 10	CLUSTER
0	0.00	1.00	2.00	0.00	0.00	3.42	3.18	2.70	1.02	2.42	1
1	1.00	2.00	2.00	0.04	1.42	3.18	3.21	2.02	1.08	2.02	1
2	1.22	0.00	0.94	0.03	0.24	1.94	3.18	2.00	1.31	2.00	0
3	0.48	1.00	0.38	0.57	0.48	1.02	3.18	2.00	1.07	2.00	0
4	0.01	1.20	1.10	0.07	1.04	2.02	3.18	2.70	1.10	2.04	1
5	0.06	1.20	0.72	0.27	0.74	1.20	3.17	2.00	1.00	2.00	0
6	0.00	1.00	0.20	0.02	0.00	1.00	3.17	2.00	1.22	2.02	0
7	0.74	1.40	0.22	0.41	0.02	1.00	3.17	2.01	1.04	2.00	0
8	1.12	1.70	1.04	0.04	0.02	2.14	3.18	2.70	1.41	2.04	1
9	0.70	1.00	0.22	0.20	1.00	1.04	3.17	2.02	2.24	3.12	0
10	1.47	1.00	0.70	0.70	1.00	2.70	3.18	2.00	1.00	2.02	0
11	0.00	2.00	0.20	0.00	0.00	2.00	3.17	2.00	1.00	3.42	0
12	0.74	1.44	2.70	0.45	0.00	1.74	3.20	2.07	1.20	2.04	1
13	0.00	1.04	2.27	0.40	1.20	1.72	3.10	2.01	2.00	2.74	1
14	0.00	1.00	2.20	0.01	1.20	2.04	3.20	2.02	2.00	2.40	1
15	1.20	0.02	1.70	0.00	1.40	2.30	3.10	2.74	1.41	2.02	1
16	0.00	1.04	1.70	0.04	0.00	1.10	3.10	2.70	1.10	2.00	1
17	0.01	1.00	2.40	0.00	1.24	1.70	3.20	2.04	1.10	2.42	1
18	0.07	1.00	1.30	0.00	0.02	3.30	3.10	2.00	1.70	2.00	1
19	0.00	1.04	2.10	0.00	1.10	1.00	3.10	2.00	1.22	2.40	1

Gambar 10. Data yang Terkelompokkan dalam Masing-masing Cluster



Gambar 9. Visualisasi Hasil Clustering

#### 4. SIMPULAN

Metode *clustering* menjadi salah satu metode *data mining* yang dapat digunakan untuk membuat pengelompokan dari suatu *datasheet*. Pengelompokan dapat digunakan untuk memetakan *datasheet* berdasar pada kategori tertentu. Penelitian yang dilakukan untuk mengelompokkan *datasheet*, hasil pengelompokan dari *datasheet* `tripadvisor_review.csv` memberikan rekomendasi banyak pengelompokan sebanyak 2 dan masing-masing kelompok diberi kategori puas, puas dan tidak puas. Proses menentukan pengelompokan sebanyak 2 berdasarkan pada evaluasi kinerja hasil clustering dengan menggunakan teknik *Davies Bouldin*, *Elbow* dan *silhouette*. Pengujian evaluasi kinerja dengan menggunakan teknik tersebut menghasilkan nilai yang sama yaitu  $K=2$ .

#### DAFTAR PUSTAKA

- [1] Morrison, *Metode Penelitian Survei*. Jakarta: Kencana, 2017.
- [2] D. Cielen, A. D. B. Meysman, and M. Ali, *Introducing Data Science*. 2016.
- [3] M. Arhami and M. Nasir, *Data Mining - Algoritma dan Implementasi*. Yogyakarta: Penerbit Andi, 2020.
- [4] M. Sholeh, R. Y. Rachmawati, and E. N. Cahyo, "Penerapan Regresi Linear Ganda Untuk Memprediksi Hasil Nilai Kuesioner Mahasiswa Dengan Menggunakan Python," vol. 11, no. 1, pp. 13–24, 2022.
- [5] C. K. Puteri and L. N. Safitri, "Analysis of linear regression on used car sales in Indonesia," *Journal of Physics: Conference Series*, vol. 1469, no. 1, 2020, doi: 10.1088/1742-6596/1469/1/012143.
- [6] B. D. F. Kurniatullah and Y. T. C. Pramudi, "Estimation of Students' Graduation Using Multiple Linear Regression Method," *Journal of Applied Intelligent System*, vol. 2, no. 1, pp. 29–36, 2017, doi: 10.33633/jais.v2i1.1415.
- [7] S. S. Rahardjo and R. Sanusi, "Linear Regression Analysis on the Determinants of Hypertension Prevention Behavior," *Journal of Health Promotion and Behavior*, vol. 4, no. 1, pp. 22–31, 2019, doi: 10.26911/thejhp.2019.04.01.03.
- [8] M. Sholeh and D. Andayati, "Machine Linear untuk Analisis Regresi Linier Biaya Asuransi Kesehatan dengan Menggunakan Python Jupyter Notebook," *Jurnal Edukasi dan Penelitian Informatika*, vol. 8, no. 1, pp. 20–27, 2022.
- [9] H. K. Pambudi, P. G. A. Kusuma, F. Yulianti, and K. A. Julian, "Prediksi Status Pengiriman Barang Menggunakan Metode Machine Learning," *Jurnal Ilmiah Teknologi Infomasi Terapan*, vol. 6, no. 2, pp. 100–109, 2020, doi: 10.33197/jitter.vol6.iss2.2020.396.
- [10] C. A. Rahardja, T. Juardi, and H. Agung, "Implementasi Algoritma K-Nearest Neighbor Pada Website Rekomendasi Laptop," pp. 75–84, 2019.
- [11] E. Sutoyo and A. Almaarif, "Educational Data Mining untuk



- Prediksi Kelulusan Mahasiswa Menggunakan Algoritme Naïve Bayes Classifier,” *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 4, no. 1, pp. 95–101, 2020, doi: 10.29207/RESTI.V4I1.1502.
- [12] Amanda and M. Veronica Sitorus, “Penerapan Algoritma K-Means Clustering Untuk Pengelompokan Konsumsi Produk Kosmetik milik PT Cedefindo,” *Jurnal Ilmiah MIKA AMIK Al Muslim*, vol. V, no. 2, pp. 63–68, 2021.
- [13] T. Hardiani, “Analisis Clustering Kasus Covid 19 Di Indonesia Menggunakan Algoritma K-Means,” *Janapati*, vol. 11, no. 2, pp. 156–165, 2022.
- [14] D. A. Manalu and G. Gunadi, “Implementasi Metode Data Mining K-Means Clustering Terhadap Data Pembayaran Transaksi Menggunakan Bahasa Pemrograman Python pada CV Digital Dimensi,” *Infotech: Journal Of Technology Information*, vol. 8, no. 1, pp. 45–54, 2022.
- [15] A. Maulana, K. Nur Akbar, and Nurahman, “Penerapan Clustering Menggunakan Algoritma K-Means Sebagai Analisis Produksi Komoditas Perikanan Provinsi di Indonesia,” *EJECTS: E-Journal Computer, Technology and Informations System*, vol. 01, no. 01, pp. 1–6, 2021.
- [16] Parjito and Permata;, “Penerapan Data Mining untuk Clustering Data Penduduk Miskin Menggunakan Metode K-Means,” *Jurnal Informatika*, vol. 3, no. 1, p. 7, 2021.
- [17] 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.
- [18] A. Badruttamam and D. A. I. Maruddani, “Penerapan Analisis Klaster K -Modes dengan Validasi Davies Bouldin Index dalam Menentukan Karakteristik Kanal Youtube di Indonesia (Studi Kasus: 250 Kanal Youtube Indonesia Teratas Menurut Socialblade ),” *JURNAL GAUSSIAN*, vol. 9, no. 3, pp. 263–272, 2020.
- [19] A. Winarta and W. J. Kurniawan, “Optimasi Cluster K-Means Menggunakan Metode Elbow pada Data Pengguna Narkoba dengan Pemrograman Python,” *Jurnal Teknik Informatika Kaputama (JTIK)*, vol. 5, no. 1, pp. 113–119, 2021.
- [20] N. T. Hartanti, “Metode Elbow dan K-Means Guna Mengukur Kesiapan Siswa SMK Dalam Ujian Nasional,” vol. 02, pp. 82–89, 2020.
- [21] I. Wahyudi *et al.*, “Analisa Penentuan Cluster Terbaik pada Metode K-Means Menggunakan Elbow Terhadap Sentra Industri Produksi di Pamekasan,” *Jurnal Aplikasi Teknologi Informasi dan Manajemen (JATIM)*, vol. 2, no. 2, pp. 72–81, 2021.
- [22] S. Paembonan and H. Abduh, “Penerapan Metode Silhouette Coefficient untuk Evaluasi Clustering Obat,” *PENA TEKNIK: Jurnal Ilmiah Ilmu-Ilmu Teknik*, vol. 6, no. 2, p. 48, 2021, doi: 10.51557/pt\_jiit.v6i2.659.
- [23] R. Hidayati *et al.*, “Analisis Silhouette Coefficient pada 6 Perhitungan Jarak K-Means Clustering,” *Techno.COM*, vol. 20, no. 2, pp. 186–197, 2021.
- [24] D. A. I. C. Dewi and D. A. K. Pramita, “Analisis Perbandingan Metode Elbow dan Silhouette pada Algoritma Clustering K-Medoids

dalam Pengelompokan Produksi Kerajinan Bali,” *Matrix: Jurnal Manajemen Teknologi dan Informatika*, vol. 9, no. 3, pp. 102–109, 2019, doi: 10.31940/matrix.v9i3.1662.

[25] A. Stewart, *Python Programming for beginners*. 2016.

[26] B. Santosa and A. Umam, *Data Mining dan Big Data Analytics*. Bantul: Penebar Media Pustaka, 2018.