PENERAPAN ALGORITMA K-MEANS PADA PENGELOMPOKAN DATA SET BAHAN PANGAN INDONESIA TAHUN 2022-2023

Alfian Adiyanto, Yudhistira Arie Wijaya
Program Studi Manajemen Informatika, STMIK IKMI Cirebon
Jl. Perjuangan No. 10B, Karyamulya Cirebon, Indonesia
Alfianadiyanto403@gmail.com

ABSTRAK

Tujuan Penelitian ini adalah melakukan pengelompok/kluster yang optimal menggunakan algoritma *k-means* pada dataset Bahan Pangan .Oleh karena itu, penelitian ini menggunakan metode analisa data Knowledge Discovery in Database (KDD) Sumber data penelitian ini diperoleh dari website Pusat Infromasi Harga Pangan Strategis Nasional. Data ini kumpulan data dari bulan Desember 2022-Januari 2023. Penelitian ini menggunakan clustering metode K-Means untuk mengelompokan data tersebut dengan metode K-Means yang sederhana. Algoritma K-Means sebuah metode pengelompokan berdasarkan jarak terdekat, dimana jarak terdekat ini digunakan untuk membagi data ke dalam sebuah cluster. Dalam penelitian ini, menghasilkan Davies Bouldin Index dari K = 2 sampai K =20, maka hasil cluster yang paling optimum adalah K = 2 dengan hasil Davies Bouldin Index sebesar 0,694.

Kata kunci : *Data mining, K-Means, Clustering, Indeks Harga Konsumen*

1. PENDAHULUAN

Data mining adalah metode untuk menemukan pola tertentu dari kumpulan data yang berjumlah besar. Meskipun banyak dipelajari pada bidang ilmu komputer dan statistika, data mining adalah metode yang bisa diterapkan dan mempermudah pekerjaan di bidang lainnya juga. Namun sebenarnya apa itu data mining?Data mining adalah metode dalam ilmu komputer yang biasa digunakan dalam proses pencarian knowledge[1]. Algoritma K-Means merupakan algoritma yang sederhana untuk diimplementrasikan, memiliki kinerja yang relatif cepat, mudah beradaptasi, dan umum digunakan[2]. Metode clustering atau pengelompokkan yang proses pemodelannya tanpa supervisi/pembelajaran serta metode pengelompokkan datanya dilakukan secara partisi. Pada metode yang digunakan dalam algoritma K-means, data akan dikelompokkan menjadi beberapa bagian kelompok dan setiap kelompok memiliki ciri-ciri yang mirip dengan satu sama lain, namun memiliki ciri-ciri yang berbeda dengan kelompok lain. Hal itu bertujuan untuk meminimalisir perbedaan antara satu cluster dan memaksimalkan perbedaan data dengan cluster lain[3].

Clustering banyak digunakan diberbagai bidang seperti biologi, psikologi, dan ekonomi. Hasil pengelompokan bervariasi karena jumlah perubahan parameter cluster maka tantangan utama analisis cluster adalah jumlah cluster atau jumlah parameter model jarang diketahui, dan harus ditentukan sebelum pengelompokan. [4]. Salah satu metode data mining yang dapat digunakan untuk memetakan atau mengelompokkan data sejenis adalah klastering. Klastering memiliki keunggulan dibandingkan metode data mining lainnya karena dapat mengklasifikasikan data tanpa pengetahuan sebelumnya. Clustering membagi data dan mengelompokkan data ke dalam cluster berdasarkan kesamaan tipe data. [5]. Dalam penentuan klasternya dapat dibuat sebanyak dua, tiga, empat dan seterusnya, dimana setiap klaster mempunyai karakteristik yang sama. Jumlah klaster terbaik dapat diperiksa menggunakan Davies-Bouldin Index (DBI). Jumlah klaster yang dipilih adalah jumlah klaster yang memiliki nilai DBI terkecil. [6]

Berdasarkan penelitian yang dilakukan oleh Viya Miralda, Muhammaf Zarlis dan Eka Irawan dengan judul "Penerapan Metode K-Means Clustering Untuk Daging Ayam Buras". Seiring dengan meningkatnya pengetahuan masyarakat dalam memilih makanan sumber protein yang baik dan sehat untuk tubuh, maka semangkin tinggi lah minat masyarakat terhadap daging ayam buras, tetapi tingginya permintaan masyarakat akan ayam buras tidak di barengi dengan produksinya, produksi ayam buras masih lebih rendah jika dibandingkan dengan ayam ras, sehingga masyarakat kesulitan menemukan daging ayam ras di pasaran. Dalam hal itu, membuat suatu penelitian dengan data mining metode clustering untuk mengelompokkan Kabupaten/Kota di Provinsi Sumatera Utara menjadi dua bagian, yang nantinya akan menghasilkan 2 kategori daerah dengan Daging ayam buras tinggi dan kategori daerah dengan Daging ayam buras rendah. Maka daerah yang termasuk dalam Daging ayam buras rendah akan lebih di tingkatkan lagi produksinya sehingga di harapkan akan bisa memenuhi permintaan akan daging ayam buras di pasar pasar yang ada di daerah kabupaten/kota di provinsi sumatera utara dan dapat memperkecil jumlah impor daging dan menambah hasil produksi ayam buras di Provinsi Sumatrea Utara, agar produksi ayam buras bisa memenuhi kebutuhan protein nasional khususnya di Provinsi Sumatera Utara.[7]

Data yang digunakan adalah Data set harga bahan pangan yang diperoleh dari website Pusat Infromasi

Harga Pangan Strategis Nasional. Data ini kumpulan data dari bulan Desember 2022-Januari 2023

Tabel 1. Sampel Dataset Bahan Pangan

No.	Provinsi	23/12/2022	…….	19/01/2023
1	Kota Medan	Rp 46.500	…….	Rp 44.400
2	Kota Pekanbaru	Rp 49.400	…….	Rp 50.650
…	….	……	…….	….
89	Kota Sorong	Rp 100.000	…….	Rp 90.000
90	Kab. Manokwari	Rp 55.000	…….	Rp 42.500

Berisi data yang berasal dari Website Bahan pangan yang setiap datanya mewakili Provini dan tanggal . Indeks Harga Konsumen (IHK) merupakan salah satu indikator untuk mengukur pertumbuhan ekonomi sehingga mengetahui fluktuasi kenaikan ataupun penurunan harga untuk mengetahui perkembangan ekonomi. Proses kenaikan harga secara general dan kontinu disebut sebagai inflasi. Salah satu indeks harga untuk mengukur inflasi antara lain IHK. [8]. Jika inflasi terjadi maka dapat mempengaruhi stabilitas keamanan dan politik. Selain itu, inflasi juga akan menyebabkan pendapatan riil masyarakat menurun, nilai tabungan turun, menyulitkan pelaku usaha untuk melakukan investasi dan produksi, angka kemiskinan meningkat dan pengangguran bertambah. Dampak yang lebih jauh lagi yaitu pertumbuhan ekonomi melambat. [9]

Tujuan dari tugas akhir ini adalah mendapatkan jumlah kluster terbaik dan untuk mengetahui anggotanya setiap cluster. Hasil pengelompokan tersebut dapat sebagai bahan pertimbangan atau analisa pemerintah dalam menganalisa harga pangan di pasar tradisional yang harganya melonjak tinggi untuk dinormalisasi harga pangannya.

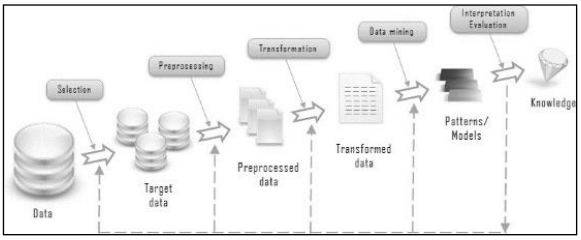
Oleh karena itu diusulkan tugas akhir dengan judul "penerapan algoritma k-means pada pengelompokan data set bahan pangan indonesia tahun 2022-2023" Adapun yang menjadi alasan dilakukannya tugas akhir ini dengan judul tersebut adalah untuk mengetahui klasterisasi dataset bahan pangan yang nanti nya diharapkan dapat membantu pemerintah mengembangkan dan meningkatkan kebijakan dibidang pembangunan ekonomi yang akan datang.

2. TINJAUAN PUSTAKA

2.1. Data Mining

Data mining adalah suatu istilah yang digunakan untuk menguraikan penemuan pengetahuan dalam database. Data mining adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan machine learning untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan terkait dari berbagai database besar. Pendekatan dasar dalam data mining adalah merangkum data dan mengambil informasi yang bermakna dan yang sebelumnya tidak diketahui

Istilah yang sering digunakan adalah pada saat melakukan pengolahan data yang benar adalah data mining dan knowledge Discovery in Database (KDD) Proses KDD sendiri digunakan untuk mendapatkan pengetahuan dari database yang ada, dimana dalam suatu database terdapat berbagai tabel-tabel yang saling berhubungan antara satu dengan yang lain. Proses KDD dan Data mining sering digunakan secara bergantian, walaupun mempunyai konsep yang berbeda tapi proses kdd dan data mining saling berkaitan. [10]. Metode penelitian yang digunakan yaitu metode *Knowledge Discovery in Database* (KDD), yang terdiri dari tahapan: *Data Selection, Preprocessing/Cleaning, Transformation Data, Data mining, dan Enterpretation/Evaluation*.[11]



Gambar 1. Tahapan proses KDD

2.2. K-Means

K-Means pertama kali dipublikasikan oleh Stuart Lloyd pada tahun 1984 dan merupakan algoritma clustering yang banyak digunakan. K-Means bekerja dengan mensegmentasi objek yang ada ke dalam kelompok atau yang disebut dengan segmen sehingga objek yang berada dalam masing-masing kelompok lebih serupa satu sama lain dibandingkan dengan objek dalam kelompok yang berbeda. [12]

K-means adalah sebuah algorima data mining unsupervised yang menggunakan centroid sebagai titik pusat dalam setiap cluster. K-means ialah algoritma yang tidak rumit, ukuran kesamaan berperan pentingdalam proses clustering menggunakan K-Means. [13] (1)

Langkah-langkah dari *K-Means (pseudocode)* :

a. Menentukan banyaknya cluster (k) untuk jumlah cluster dari dataset yang ada
b. Melnelntulkan k selbagai Celntroid, biasanya dilakulkan selcara acak (random).
c. Hitulng jarak data delngan celntroid melnggulnakan rulmuls jarak melnggulnakan rulmuls Elulclidelan (pelrsamaan 1)
d. Distancel: $d(x\text{i}, \mu\text{j}) = \sqrt{\sum}(x\text{i}, \mu\text{j})$
e. dimana d melrulpakan titik dokulmeln, xi melrulpakan data kritelria dan µj melrulpakan celntroid pada clulstelr kel-j.
f. Kellompokkan data belrdasarkan keldelkatan delngan celntroid kelmuldian pelrbaharuli nilai celntroid barul delngan lokasi dari pulsat clulstelr melnggulnakan pelrsamaan 2
g. Lakulkan langkah 2 sampai 4 sampai anggota

tiap cluster tidak ada yang berubah.(Indriyani & Irfiani, 2019)

2.3. Clustering

Clustering adalah algoritma pengelompokkan data dengan jumlah besar di proses menjadi bagian kecil dengan atribut kemiripan tertentu**.** K-Means Clustering termasuk pada metode penganalisaan data atau metode data mining yang proses nya tanpa pengawasan (unsupervised) serta menjadi suatu metode pengelompokan yang datanya dilakukan dengan sistem partisi. [15]. Pada clustering partisi dari objek data yang mempunyai karakteristik sama akan dikelompokkan padasatu kelompok dan data yang memiliki karakteristik berbeda akan dikelompokkan pada kelompok yang lainnya.[16]

3. METODE PENELITIAN

3.1. Sumber Data

Sumber data dalam penelitian ini bersumber dari Website https://hargapangan.id/. Data diambil pada hari Kamis,19 Januari 2023,Pada Pukul 17.00 WIB. Data memiliki 2 atribut yaitu Provinsi dan Tanggal.Data yang digunakan selama 23 Desember 2022- 19 Januari 2023.

3.2. Tekhnik Pengumpulan Data

Data penelitian terbagi menjadi dua data yaitu data primer dan data sekunder. Data primer merupakan data yang diperoleh peneliti secara langsung, sementara data sekunder adalah data yang diperoleh peneliti dari sumber yang sudah ada. Contoh data primer adalah data yang diperoleh dari responden melalui kuesioner, wawancara peneliti dengan narasumber, panel, kelompok fokus. Contoh data sekunder misalnya catatan atau dokumentasi perusahaan berupa absensi, gaji, laporan keuangan publikasi perusahaan, laporan pemerintah, data yang diperoleh dari majalah, dan lain sebagainya Pada pengumpulan data ini, menggunakan teknik pengumpulan data sekunder untuk mendapatkan hasil yang akurat. Data yang digunakan merupakan data yang sudah jadi yang didapatkan dari Website Bahan Pangan https://hargapangan.id/.

3.3. Tahapan Perancangan

Pada tahap perancangan atau penelitian ini, metode yang akan digunakan adalah metode Knowledge Discovery in Databases (KDD), Adapun tahapan sebagai berikut :

a. Data Selection
   Menyaring sekumpulan data mentah untuk menentukan variabel data apa saja yang akan diproses.
b. Preprocessing
   Prepocessing merupakan tahap membersihkan data dari noise, missing valuedan data yang tidak relevan.
c. Transformation
   Mentransformasikan/ merubah data ke dalam bentuk yang sesuai dengan format pengolahan data pada algoritma data mining yang akandigunakan. Setiap algoritma data mining mengolah data dengan format tertentu.
d. Data Mining
   Merupakan tahap inti dari KDD yaitu mengekstraksi sekumpulan data dengan menggunakan metode dan algoritma sehingga ditemukan pengetahuan.
e. Interpretation/Evaluation
   Menginterpretasikan pengetahuan atau pola yang ditemukan dari proses data mining serta mengevaluasi apakah pola tersebut sudah sesuai dengan target yang diinginkan. [17]

4. HASIL DAN PEMBAHASAN

4.1. Pengelompokan Dataset menggunakan Algoritma K Means Clustering

Penerapan Algoritma K-Means Clustering dirancang oleh Tools Rapidminer. Pada penelitian ini, metode yang akan digunakan adalah metode Knowledge Discovery in Databases (KDD. Penerapan KDD sebagai berikut :

4.2. *Data Selection*

Read Excel, Operator *Read Excel* yang berfungsi sebagai pembaca *file excel*. Read Excel merupakan operator yang paling dasar digunakan sebelum memulai sebeuah proses. Operator ini dapat digunakan untuk memuat data dari spreadsheet Microsoft Excel. Data set yang digunakan adalah dataset bahan pangan.

Gambar 2. Read Excel

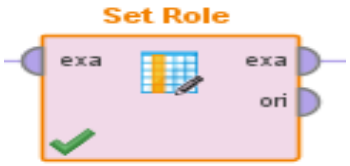
Parameter pada operator *Read Excel* menggunakan Parameter *default.* Dari hasil pembacaan operator *Read Excel* didapat infomasi sebagai berikut :

Tabel 2. Statistik Dataset

No	Uraian	Keterangan
1.	Record	90
2.	Special Attribute	0
3.	Regular Attribute	22
4.	Attribute :	
	Provinsi	Nominal, missing 0
	23/12/2022	Integer, missing 0
	26/12/2022	Integer, missing 0
	27/12/2022	Integer, missing 0
	28/12/2022	Integer, missing 0
	29/12/2022	Integer, missing 0
	30/12/2022	Integer, missing 0
	02/01/2023	Integer, missing 0
	03/01/2023	Integer, missing 0

No	Uraian	Keterangan
	04/01/2023	Integer, missing 0
	05/01/2023	Integer, missing 0
	06/01/2023	Integer, missing 0
	09/01/2023	Integer, missing 0
	10/01/2023	Integer, missing 0
	11/01/2023	Integer, missing 0
	12/01/2023	Integer, missing 0
	13/01/2023	Integer, missing 0
	16/01/2023	Integer, missing 0
	17/01/2023	Integer, missing 0
	18/01/2023	Integer, missing 0
	19/01/2023	Integer, missing 0

Set Role, Operator *Set Role* berfungsi digunakan untuk mengubah peran satu atau lebih Atribut.



Gambar 3. Set Role

Parameter pada operator *Set Role* menggunakan Parameter *default.* Dari hasil pembacaan operator *Set Role* didapat infomasi sebagai berikut :

Tabel 3. Statistik dataset setelah menggunakan Set Role

No	Uraian	Keterangan
1.	Record	90
2.	Special Attribute	2
3.	Regular Attribute	20
4.	Attribute :	
	Cluster	Nominal, missing 0
	23/12/2022	Integer, missing 0
	26/12/2022	Integer, missing 0
	27/12/2022	Integer, missing 0
	28/12/2022	Integer, missing 0
	29/12/2022	Integer, missing 0
	30/12/2022	Integer, missing 0
	02/01/2023	Integer, missing 0
	03/01/2023	Integer, missing 0
	04/01/2023	Integer, missing 0
	05/01/2023	Integer, missing 0
	06/01/2023	Integer, missing 0
	09/01/2023	Integer, missing 0
	10/01/2023	Integer, missing 0
	11/01/2023	Integer, missing 0
	12/01/2023	Integer, missing 0
	13/01/2023	Integer, missing 0
	16/01/2023	Integer, missing 0
	17/01/2023	Integer, missing 0
	18/01/2023	Integer, missing 0
	19/01/2023	Integer, missing 0

Select Attributes, Operator attributes berfungsi memilih subset dari Atribut dari ContohSet dan menghapus Atribut lainnya.

Gambar 4. Select Atributes

Parameter pada operator *Select Attributes* yang digunakan tampak pada tabel dibawah ini.

Tabel 4. Parameter dan atribut yang dipilih pada operator Select Attributes

No	***parameter***	***isi***
1.	*Attribute Filter Type*	*Subset*
2.	*Selected Attributes*	No, 23/12/2022, 26/12/2022, 27/12/2022,28/12/2022, 29/12/2022,30/12/2022, 02/01/2023,03/01/2023, 04/01/2023,05/01/2023, 06/01/2023,09/01/2023, 10/01/2023,11/01/2023, 12/01/2023,13/01/2023, 16/01/2023,17/01/2023, 18/01/2023, 19/01/2023

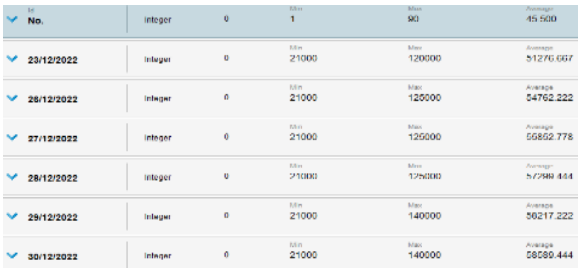
Dari hasil pembacaan operator *Select Attribut* didapat informasi sebagai berikut.

Tabel 5. Statistik Dataset

No	Uraian	Keterangan
1.	Record	90
2.	Special Attribute	0
3.	Regular Attribute	22
4.	Attribute :	
	Provinsi	Nominal, missing 0
	23/12/2022	Integer, missing 0
	26/12/2022	Integer, missing 0
	27/12/2022	Integer, missing 0
	28/12/2022	Integer, missing 0
	29/12/2022	Integer, missing 0
	30/12/2022	Integer, missing 0
	02/01/2023	Integer, missing 0
	03/01/2023	Integer, missing 0
	04/01/2023	Integer, missing 0
	05/01/2023	Integer, missing 0
	06/01/2023	Integer, missing 0
	09/01/2023	Integer, missing 0
	10/01/2023	Integer, missing 0
	11/01/2023	Integer, missing 0
	12/01/2023	Integer, missing 0
	13/01/2023	Integer, missing 0
	16/01/2023	Integer, missing 0
	17/01/2023	Integer, missing 0
	18/01/2023	Integer, missing 0
	19/01/2023	Integer, missing 0

4.3. *PreProcessing*

Dikarenakan pada dataset Bahan Pangan tidak ditemukan *missing value* atau data yang tidak memiliki nilai maka *preprocessing* tidak dilakukan.



Gambar 5. Hasil Statistik

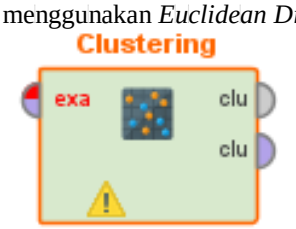
Setelah dilihat dari hasil statistik, tidak ditemukan adanya *missing value*. Maka proses dapat dilanjutkan ke tahap selanjutnya.

4.4. *Data Transformation*

Data Transformasi merupakan langkah untuk Mentransformasikan atau merubah data ke dalam bentuk yang sesuai dengan format pengolahan data pada algoritma data mining yang akandigunakan. Pada langkah ini tidak dilakukan karena tipe dataset yang digunakan sudah bertipe numerik atau angka seperti pada gambar 5.

4.5. *Data Mining*

Operator ini melakukan *clustering* menggunakan *algoritma k-means*. Pada tahap ini bagian *parameter* menggunakan *measure types Numerical Measure* penelitian ini menggunakan *Euclidean Distance*.



Gambar 6. K-Means Clustering

Parameter pada operator *K-Means Clustering* yang digunakan tampak pada tabel dibawah ini. Parameter pada operator *K-Means Clustering* ini menggunakan parameter default.

Tabel 6. Parameter pada operator K-Means Clustering

No	*Parameter*	Isi
1.	*K*	2-20

Dari hasil pembacaan operator *K-Means Clustering* dengan parameter *default* didapat informasi sebagai berikut.

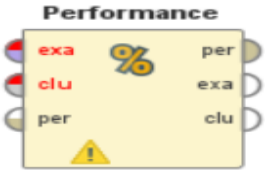
Tabel 7. Hasil K-Means Clustering pada k=2

No	*Cluster*	Keterangan
1.	*Cluster 0*	74
2.	*Cluster 1*	16
	Total Number of Items	90

Tabel 8. Hasil K-Means Clustering pada k=20

No	*Cluster*	Ktrgan
1.	*Cluster 0*	5
2.	*Cluster 1*	11
3.	*Cluster 2*	1
4.	*Cluster 3*	4
5.	*Cluster 4*	6
6.	*Cluster 5*	7
7.	*Cluster 6*	2
8.	*Cluster 7*	2
9.	*Cluster 8*	7
10.	*Cluster 9*	7
11.	*Cluster 10*	1
12.	*Cluster 11*	1
13.	*Cluster 12*	2
14.	*Cluster 13*	14
15.	*Cluster 14*	3
16.	*Cluster 15*	3
17.	*Cluster 16*	9
18.	*Cluster 17*	1
19.	*Cluster 18*	1
20.	*Cluster 19*	3
	Total Number of Items	90

Selanjutnya digunakan opereator *Performance* dengan metode *Davies-Bouldin Index (DBI)*. Untuk mengetahui nilai *DBI* nya.



Gambar 7. Cluster Distance Performance

Parameter pada operator *Performance* yang digunakan tampak pada tabel dibawah ini.

Tabel 9. Parameter pada operator Performance

No	*Parameter*	Isi
1.	*Main Criterion*	Davies Bouldin

Dari hasil pembacaan operator *Performance* didapat informasi sebagai berikut.

Tabel 10. Hasil operator Performance (DBI)

No	K	*DBI*
1.	2	0,694
2.	3	0,821
3.	4	0,782
4.	5	0,889
5.	6	0,875
6.	7	0,727
7.	8	0,909
8.	9	1,111
9.	10	0,956
10.	11	0,998
11.	12	1,037
12.	13	0,992
13.	14	0,920
14.	15	1,011

No	K	*DBI*
15.	16	0,006
16.	17	0,954
17.	18	0,946
18.	19	0,891
19.	20	0,907

4.6. *Interpretation/Evaluation*

Setelah melakukan perbandingan *DBI* dengan metode *K-Means* dari k-2 sampai k-20 yang terdapat pada tabel diatas, dapat dilihat bahwa *cluster* terkecil yang mendekati 0 yaitu k-2, dengan nilai *DBI* sebesar 0,694 Karena nilai k-2 merupakan nilai terkecil dibandingkan k lainnya, maka dapat disimpulkan bahwa k-2 dengan nilai 0,694 yang paling mendekati 0 merupakan hasil *cluster* terbaik. Dengan jumlah *cluster* 0 sebanyak 74 *items*, dan *cluster* 1 sebanyak 16 *items*.

5. KESIMPULAN DAN SARAN

Berdasarkan pengelompokkan data yang sudah dilakukan dengan metode *K-Means* menggunakan parameter *default* diperoleh hasil terkecil yaitu pada *cluster* k-2 dengan nilai *DBI* sebesar 0,694. Maka *cluster* k-2 merupakan *cluster* terbaik dari total k sebanyak 20. Dengan jumlah *cluster* 0 sebanyak 74 *items*, dan *cluster* 1 sebanyak 16 *items.*

Untuk peneliti selanjutnya bisa menggunakan metode /algoritma yang lain seperti klasifikasi, Algoritma Random Forest ,Naïve Bayes, agar mendapatkan hasil yang lebih bagus lagi baik dari nilai DBI maupun kelompok.

DAFTAR PUSTAKA

[1] H. Prastiwi, J. Pricilia, and E. Raswir, "Implementasi Data Mining Untuk Menentuksn Persediaan Stok Barang Di Mini Market Menggunakan Metode K-Means Clustering Jurnal Informatika Dan Rekayasa Komputer (JAKAKOM)," *J. Inform. Dan Rekayasa Komput.*, vol. 1, no. April, pp. 141–148, 2022.

[2] S. N. Br Sembiring, H. Winata, and S. Kusnasari, "Pengelompokan Prestasi Siswa Menggunakan Algoritma K-Means," *J. Sist. Inf. Triguna Dharma (JURSI TGD)*, vol. 1, no. 1, p. 31, 2022, doi: 10.53513/jursi.v1i1.4784.

[3] F. N. Dhewayani, D. Amelia, D. N. Alifah, B. N. Sari, and M. Jajuli, "Implementasi K-Means Clustering untuk Pengelompokkan Daerah Rawan Bencana Kebakaran Menggunakan Model CRISP-DM," *J. Teknol. dan Inf.*, vol. 12, no. 1, pp. 64–77, 2022, doi: 10.34010/jati.v12i1.6674.

[4] S. Handoko, F. Fauziah, and E. T. E. Handayani, "Implementasi Data Mining Untuk Menentukan Tingkat Penjualan Paket Data Telkomsel Menggunakan Metode K-Means Clustering," *J. Ilm. Teknol. dan Rekayasa*, vol. 25, no. 1, pp. 76–88, 2020, doi: 10.35760/tr.2020.v25i1.2677.

[5] L. M. Harahap, W. Fuadi, L. Rosnita, E. Darnila, and R. Meiyanti, "Klastering Sayuran Unggulan Menggunakan Algoritma K-Means," *J. Tek. Inform. dan Sist. Inf.*, vol. 8, no. 3, pp. 567–579, 2022, doi: 10.28932/jutisi.v8i3.5277.

[6] A. Badruttamam, S. Sudarno, 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)," *J. Gaussian*, vol. 9, no. 3, pp. 263–272, 2020, doi: 10.14710/j.gauss.v9i3.28907.

[7] V. Miralda, M. Zarlis, and E. Irawan, "Penerapan Metode K-Means Clustering Untuk Daging Ayam Buras," *Build. Informatics, Technol. Sci.*, vol. 2, no. 2, pp. 91–98, 2020, doi: 10.47065/bits.v2i2.493.

[8] F. Azizah and M. Athoillah, "Analisis Dampak Covid-19 Terhadap Indeks Harga Konsumen dengan K-Means dan Regresi Berganda," *Indones. J. Appl. Stat.*, vol. 4, no. 1, p. 21, 2021, doi: 10.13057/ijas.v4i1.46329.

[9] S. U. Wijaya and N. N. Ngatini, "Pengembangan Pemodelan Harga Beras di Wilayah Indonesia Bagian Barat dengan Pendekatan Clustering Time Series," *Limits J. Math. Its Appl.*, vol. 17, no. 1, p. 51, 2020, doi: 10.12962/limits.v17i1.5994.

[10] M. Intan Pratiwi Hant and Hendry, "Data Mining Technique Using Naïve Bayes Algorithm To Predict Shopee Consumer Satisfaction Among Millennial Generation," *J. Tek. Inform.*, vol. 3, no. 4, pp. 829–838, 2022, [Online]. Available: https://doi.org/10.20884/1.jutif.2022.3.4.295.

[11] N. L. P. P. Dewi, I. N. Purnama, and N. W. Utami, "Penerapan Data Mining Untuk Clustering Penilaian Kinerja Dosen Menggunakan Algoritma K-Means (Studi Kasus: STMIK Primakara)," *J. Ilm. Teknol. Inf. Asia*, vol. 16, no. 2, p. 105, 2022, doi: 10.32815/jitika.v16i2.761.

[12] A. Aditya, I. Jovian, and B. N. Sari, "Implementasi K-Means Clustering Ujian Nasional Sekolah Menengah Pertama di Indonesia Tahun 2018/2019," *J. Media Inform. Budidarma*, vol. 4, no. 1, p. 51, 2020, doi: 10.30865/mib.v4i1.1784.

[13] R. A. Farissa, R. Mayasari, and Y. Umaidah, "Perbandingan Algoritma K-Means dan K-Medoids Untuk Pengelompokkan Data Obat dengan Silhouette Coefficient di Puskesmas Karangsambung," *J. Appl. Informatics Comput.*, vol. 5, no. 2, pp. 109–116, 2021, doi: 10.30871/jaic.v5i1.3237.

[14] F. Indriyani and E. Irfiani, "Clustering Data

Penjualan pada Toko Perlengkapan Outdoor Menggunakan Metode K-Means," *JUITA J. Inform.*, vol. 7, no. 2, p. 109, 2019, doi: 10.30595/juita.v7i2.5529.

[15] D. Isya Ramadhani, O. Damayanti, O. Thaushiyah, and A. Rahman Kadafi, "Penerapan Metode K-Means Untuk Clustering Desa Rawan Bencana Berdasarkan Data Kejadian Terjadinya Bencana Alam," *J. Ris. Komputer)*, vol. 9, no. 3, pp. 2407–389, 2022, doi: 10.30865/jurikom.v9i3.4326.

[16] G. Gustientiedina, M. H. Adiya, and Y. Desnelita, "Penerapan Algoritma K-Means Untuk Clustering Data Obat-Obatan," *J. Nas. Teknol. dan Sist. Inf.*, vol. 5, no. 1, pp. 17–24, 2019, doi: 10.25077/teknosi.v5i1.2019.17-24.

[17] H. Syahputra, "Clustering Tingkat Penjualan Menu (Food and Beverage) Menggunakan Algoritma K-Means," *J. KomtekInfo*, vol. 9, pp. 29–33, 2022, doi: 10.35134/komtekinfo.v9i1.274.