

Stock Data Clustering of Food and Beverage Company

Shofwatul 'Uyun¹, Subanar²

¹Computer Science Graduate Program, Gadjah Mada University, Yogyakarta

²Faculty of Mathematics and Natural Science, Gadjah Mada University, Yogyakarta

Abstract

Cluster analysis can be defined as identifying groups of similar objects to discover distribution of patterns and interesting correlations in large data sets. Clustering analysis is important in the fields of pattern recognition and pattern classification. Over the years many methods have been developed for clustering data. In general, clustering methods can be categorized into two categories, i.e., fuzzy clustering and hard clustering. Fuzzy C-means is one of many methods of clustering based on fuzzy approach, while K-Means and K-Medoid are methods clustering based on crisp approach.

This study aims to apply Fuzzy C-Means, K-Means and K-Medoid methods for clustering stock data in a food and beverage company. The main goal is to find a clustering method that can produce optimal clusters. The resulting clusters are validated using Dunn's Index (DI). It is expected that the result of this research can be used to support decision making in the food and beverage company.

Keywords : Clustering, Fuzzy C-Means, K-Means, K-Medoid, Cluster Validity, Dunn's Index (DI)

1. Pengantar

Sampai saat ini perkembangan teknologi lebih banyak berfokus dan berkonsentrasi pada pengumpulan data dan juga penyimpanannya, tetapi tidak memikirkan pengelolaan dan pemanfaatan data yang telah dikumpulkan tersebut. Hal ini menunjukkan bahwa informasi merupakan jantung dari operasional bisnis, selain itu para manajer dapat memanfaatkan data yang disimpan menjadi pengetahuan (*knowledge*) yang sangat berharga dalam dunia bisnis. Disinilah letak dimana *clustering* dapat membantu pengelompokan data saham dari beberapa perusahaan.

Clustering sangat dibutuhkan dalam melakukan analisis pasar modal. Seperti halnya bisnis-bisnis lainnya, bahwa data pada pasar modal sangatlah besar dan juga memerlukan proses yang sangat cepat. Untuk menghilangkan kekaburan pada data besar, salah satu metode yang umum digunakan adalah *clustering*.

Pengklusteran data merupakan suatu cara untuk mencari kesamaan pada data dan menempatkan data yang serupa tersebut menjadi beberapa kelompok. Pembagian

pengklusteran dari sekumpulan data menjadi beberapa kelompok jika kesamaannya dalam sebuah kelompok tersebut lebih besar dibandingkan dengan kelompok yang lain. Ide dari pengelompokan atau *clustering* ini sebenarnya sederhana tanpa memerlukan bantuan dari pemikiran manusia terlebih dahulu, yaitu jika sewaktu-waktu ada data dalam jumlah yang besar dan biasanya dengan data yang besar tersebut, *clustering* digunakan untuk meringkas data yang sangat besar tersebut menjadi beberapa kelompok yang kecil atau beberapa kategori perintah untuk memudahkan analisis selanjutnya.

2. Tinjauan Pustaka

2.1 Clustering

Klasifikasi kelompok data tak disupervisi, seperti persoalan berbagai disiplin ilmu dapat diselesaikan dengan sebuah metode analisis data yang dikenal sebagai *clustering*. *Clustering* akan menjadi lebih berguna dengan lebih banyak data. Suatu adopsi secara luas untuk definisi dari *clustering* yang optimal adalah suatu

pemartisian yang meminimalkan jarak dalam *cluster* dan memaksimalkan jarak antar *cluster*.

Metode untuk mendapatkan *cluster* kumpulan data dapat dibagi menjadi dua, yaitu *hierarchical* dan *non-hierarchical*. Tujuan analisis *cluster* adalah untuk membentuk suatu kumpulan obyek kedalam *cluster* sedemikian hingga obyek-obyek dengan *cluster* yang sama mempunyai tingkat derajat *similaritas*, sedangkan obyek-obyek yang mempunyai *cluster* yang berbeda mempunyai tingkat derajat *dissimilaritas*. Selain itu, tujuan lain dari analisis *cluster* adalah untuk mengklasifikasikan beberapa objek berdasarkan kemiripan diantara beberapa objek tersebut dan mengorganisasi data tersebut menjadi beberapa *group*.

2.2 Fuzzy C-Means (FCM)

Apabila dimiliki sampel n objek yang masing-masing dicirikan karakteristiknya dengan p atribut (*feature*) dan terhimpun dalam $X = \{x_1, x_2, \dots, x_n\}$ dengan setiap $x_k \in \mathcal{R}^p$ $1 \leq k \leq n$, maka proses *clustering* adalah upaya pengelompokan X menjadi c *cluster* sedemikian sehingga jarak antar objek dalam *cluster* yang sama relatif lebih kecil dari pada jarak antar *cluster*. Jika titik pusat *cluster* (*cluster center*) atau *cluster prototype* ke- i dinyatakan dalam v_i , maka jarak objek x_j terhadap pusat *cluster* v_i ditentukan dengan rumus jarak. Beberapa rumusan jarak antara lain :

- Jarak Euclidean**

$$d^2(x_j, v_i) = (x_j - v_i)^T (x_j - v_i) \quad (1)$$

- Jarak Minskowski**

$$d(x_k, v_i) = \left[\sum_{j=1}^p |x_{kj} - v_{ij}|^m \right]^{1/m} \quad (2)$$

- Jarak Manhattan**

$$d(x_k, v_i) = \left[\sum_{j=1}^p |x_{kj} - v_{ij}| \right] \quad (3)$$

Jarak Manhattan adalah Jarak Minskowski dengan $m=1$

Clustering dengan pendekatan tegas setiap objek hanya dapat menjadi anggota dari satu *cluster* saja sedangkan pada

pendekatan *fuzzy* mengasumsikan bahwa objek ke- k ($k=1,2,\dots,n$) dianggap menjadi anggota dari semua *cluster* ke- i ($i=1,2,\dots,c$) dengan tingkat keanggotaan antara 0 sampai 1. *Fuzzy clustering* pertama kali dikembangkan oleh [1]. Pengembangan dengan penerapan metode *least-square* dibangun oleh [2], yang diformulasikan lebih general oleh Bezdek [3] menjadi algoritma yang terkenal dengan sebutan *Fuzzy-C-Means Clustering* (FCM). Selanjutnya FCM mengalami berbagai modifikasi dan pengembangan oleh banyak peneliti.

Konsep FCM merupakan ide mencari struktur *cluster* yang meminimalkan jarak antar objek ke setiap titik pusat *cluster* (*sum-square error within group*). Bentuk fungsi tujuan yang paling klasik dari FCM adalah :

$$J_1(U, V; X) = \sum_{i=1}^c \sum_{k=1}^n \mu_{ik} d_{ik}^2 (v_i - x_k) \quad (4)$$

dimana $V = \{v_1, v_2, \dots, v_c\}$, $v_i \in \mathcal{R}^p$ adalah vektor pusat-pusat *cluster*, dan U adalah matrik keanggotaan atas X . Partisi optimal U^* untuk X adalah pasangan (U^*, V^*) yang merupakan *local minimizer* untuk J_1 .

Dunn [2] menggeneralisir bentuk J_1 menjadi bentuk J_m dengan $m=2$, yaitu :

$$J_m(U, V; X) = \sum_{i=1}^c \sum_{k=1}^n \mu_{ik}^m d_{ik}^2 (v_i - x_k) \quad (5)$$

Sedangkan Bezdek [4], merumuskan bentuk (5) ke bentuk yang lebih luas dengan batasan $m \in (1, \infty)$ dan generalisasi untuk rumus jarak menjadi persamaan :

$$J_m(U, V, X) = \sum_{i=1}^c \sum_{k=1}^n \mu_{ik}^m (v_i - x_k)^T A (v_i - x_k) \quad (6)$$

dengan A adalah sembarang matrik definit positif. Nilai m yang baik umumnya diambil antara 1.5 sampai 4, tetapi yang umum diambil adalah 2 [5].

Bila M_{cn} adalah himpunan seluruh matrik $cn \times n$, yaitu matrik partisi yang mungkin, maka $U \in M_{cn}$ sehingga optimasi fungsi (6) merupakan optimasi dengan tujuan meminimalkan J_m dengan kendala :

$$1. \mu_{ik} \in [0,1] \text{ untuk: } 1 \leq i \leq c \text{ dan } 1 \leq k \leq n \quad (7)$$

$$2. \sum_{i=1}^c \mu_{ik} = 1 \text{ untuk: } 1 \leq k \leq n \quad (8)$$

$$3. 0 < \sum_{k=1}^n \mu_{ik} < n \text{ untuk: } 1 \leq i \leq c \quad (9)$$

Minimasi (6) dengan kendala (7), (8) dan (9) dilakukan dengan mendefereensialkan J_m terhadap v_i (untuk matrik U dianggap fix) dan terhadap μ_{ik} (untuk v dianggap tetap) [3] mendapatkan nilai :

$$\mu_{ik} = \frac{\left[\frac{1}{\sum_{j=1}^c \frac{1}{\|x_k - v_j\|^2} \right]^{1/(m-1)}}{\left[\frac{1}{\sum_{j=1}^c \frac{1}{\|x_k - v_j\|^2} \right]^{1/(m-1)}} \quad (10)$$

untuk $i=1,2,...,c; k=1,2,...,n$
dan :

$$v_i = \frac{1}{\sum_{k=1}^n \mu_{ik}^m} \sum_{k=1}^n \mu_{ik}^m x_k \quad (11)$$

untuk $i=1,2,...,c$

Selanjutnya algoritma FCM dapat dituliskan sebagai berikut:

Algoritma 1. FCM

Step 0 : Input : matriks X , m , c , ϵ , α (iterasi)

Step 1 : Inisialisasi $U^{(0)}$ dan $V^{(0)}$ secara random $[0,1]$

Repeat

Step 2 : Update $V^{(\alpha)}$ dengan persamaan (11)

Step 3 : Update $U^{(\alpha)}$ dengan persamaan (10)

Until $\max \|U^{(\alpha)} - U^{(\alpha-1)}\| < \epsilon$

2.3 K-Means Clustering

Beberapa metode *hard clustering* merupakan metode yang sederhana dan populer. Dari sekumpulan data yang memiliki dimensi $N \times n$ yang akan diselesaikan dengan metode K-means yang

mengalokasikan beberapa data kebeberapa cluster c untuk meminimalkan *sum of squares*. Untuk mencari *distance norm* digunakan persamaan berikut :

$$D_{ik}^2 = (x_k - v_i)^T (x_k - v_i), \quad (12)$$

$1 \leq i \leq c, 1 \leq k \leq N$

Pada K-Means clustering v_i disebut dengan beberapa prototype cluster. Pusat cluster dapat dihitung dengan persamaan (13) :

$$v_i = \frac{\sum_{k=1}^{N_i} x_k}{N_i}, x_k \in A_i, \quad (13)$$

Dimana, N_i adalah banyaknya objek pada A_i .

Algoritma 2. K-Means

Step 0 : Input : matriks X , c , α (iterasi)

Step 1 : Inisialisasi $V^{(0)}$ secara random dari matriks X

Repeat

Step 2 : Compute D_{ik}^2 dengan persamaan (12)

Step 3 : Select the points for a cluster with the minimal D_{ik}^2

Step 4 : calculate $V^{(\alpha)}$ dengan persamaan (13)

Until $\max \|V^{(\alpha)} - V^{(\alpha-1)}\| \neq 0$

Ending Calculate the partition matrix

2.4 K-Medoid Clustering

K-Medoid clustering merupakan metode *hard clustering*. Dari sekumpulan data yang memiliki dimensi $N \times n$ yang akan diselesaikan dengan metode *K-Medoid* yang mengalokasikan beberapa data ke beberapa cluster c untuk meminimalkan *sum of squares*:

$$\sum_{i=1}^c \sum_{k \in A_i} \|x_k - v_i\|^2 \quad (14)$$

Dimana, A_i adalah sekumpulan objek (beberapa point data) pada cluster ke- i dan v_i adalah rata-rata untuk beberapa cluster i .

Pada persamaan (14) merupakan *distance norm*. Pada K-Medoid clustering, pusat cluster adalah beberapa objek yang paling dekat pada rata-rata data pada satu cluster $V = \{v_i \in X / 1 \leq i \leq c\}$

Algoritma 3. K-Medoid

Step 0 : Input : matriks X , c , α (iterasi)

Step 1 : Inisialisasi $V^{(0)}$ secara random dari matriks X

Repeat

Step 2: Compute D_{ik}^2 dengan persamaan (12)

Step 3: Select the points for a cluster with the minimal D_{ik}^2

Step 4: Calculate cluster centers (the original K-means):

$$v_i^{(l)} = \frac{\sum_{j=1}^{N_i} x_j}{N_i}$$

Step 5: Choose the nearest data point to be the cluster center

$$D_{ik}^{2*} = (x_k - v_i^*)^T (x_k - v_i^*)$$

$$x_i^* = \arg \min_i (D_{ik}^{2*}); v_i^{(l)} = x_i^*$$

Until $\max \|V^{(\alpha)} - V^{(\alpha-1)}\| \neq 0$

Ending Calculate the partition matrix

2.5 Validitas Clustering

Menentukan jumlah cluster merupakan suatu hal yang penting untuk memperoleh hasil yang berarti. Hasil *clustering* tidak akan dapat menggambarkan struktur data yang real jika jumlah cluster yang dipakai tidak tepat. Oleh karena itu, menemukan validitas *clustering* merupakan hal utama dalam analisis cluster. Validitas cluster dapat digunakan untuk menentukan berapa sebaiknya nilai c , sehingga hasil *clustering* dapat dianggap terbaik. Ukuran ini dikenal sebagai validitas *clustering*. Pada penelitian ini metode yang digunakan untuk mencari

validitas *clustering* adalah *Dunn's Index* (DI)

Dunn's index merupakan fungsi validitas yang mampu memberikan hasil penilaian yang efektif untuk aplikasi yang menggunakan beberapa metode *clustering* yang berbeda [6]. Index ini digunakan berdasarkan pada ide pengidentifikasian beberapa cluster yang padat (*compact*) dan dapat terpisah dengan baik antara satu cluster dengan yang lainnya. *Dunn's Index* didefinisikan sebagai berikut :

$$DI(c) = \min_{1 \leq i \leq c} \left\{ \min_{\substack{1 \leq j \leq c \\ j \neq i}} \left\{ \frac{\delta(X_i, X_j)}{\max_{1 \leq k \leq c} \{\Delta(X_k)\}} \right\} \right\} \quad (15)$$

$\delta(X_i, X_j)$ adalah jarak antar cluster X_i dan X_j (*intercluster distance*); $\Delta(X_k)$ merupakan *intracluster distance* dari cluster X_k ; dan c adalah banyaknya jumlah cluster. Tujuan utama pengukuran ini adalah untuk memaksimalkan jumlah jarak *intracluster*. Nilai dari DI yang paling besar menunjukkan cluster yang baik.

3. Cara Penelitian

3.1. Bahan Penelitian

- Data saham perusahaan makanan dan minuman sebagai bahan penelitian utama
- Pustaka-pustaka untuk melakukan proses *clustering*.

3.2. Peralatan Software

- Sistem Operasi *Windows XP Professional Service Pack 2*.
- *Matlab 7.0*

3.3. Peralatan Hardware

- *CPU* dengan *Processor Intel Pentium III 600 MHz*
- *Memory DDRAM 128 MB*
- *HardDisk 20GB*
- *Display Adapter VGA 8 MB*
- *Monitor* dengan resolusi 1024×768 pixel

4. Hasil dan Pembahasan

4.1. Data Saham

Pada penelitian ini data yang digunakan adalah data saham dari perusahaan makanan dan minuman yang dikelompokkan berdasarkan produk yang dihasilkan. Ada 7 kelompok kategori perusahaan makanan dan minuman :

Tabel 4.1. Kode dan Nama Emiten Perusahaan Makanan dan Minuman yang Memproduksi Minuman dalam Kemasan

Kode Emiten	Nama Emiten
ADES	Ades Alfindo Tbk
AQUA	Aqua Golden Mississippi
SUBA	Suba Indah Tbk

Tabel 4.2. Kode dan Nama Emiten Perusahaan Makanan dan Minuman yang Memproduksi Coklat Bubuk

Kode Emiten	Nama Emiten
DAVO	Davomas Abadi Tbk
PSDN	Prashida Aneka Niaga Tbk
CEKA	Cahaya Kalbar Tbk

Tabel 4.3. Kode dan Nama Emiten Perusahaan Makanan dan Minuman yang Memproduksi Minuman Beralkohol

Kode Emiten	Nama Emiten
DLTA	Delta Djakarta Tbk
MLBI	Multi Bintang Indonesia Tbk

Tabel 4.4. Kode dan Nama Emiten Perusahaan Makanan dan Minuman yang Memproduksi Mie Instant

Kode Emiten	Nama Emiten
INDF	Indofood Sukses Makmur Tbk
SKLT	Sekar Laut Tbk
STTP	Siantar Top Tbk
AISA	Asia Intiselera

Tabel 4.5. Kode dan Nama Emiten Perusahaan Makanan dan Minuman yang Memproduksi Biskuit

Kode Emiten	Nama Emiten
MYOR	Mayora Indah Tbk
SIPD	Sierad Produce Tbk

Tabel 4.6. Kode dan Nama Emiten Perusahaan Makanan dan Minuman yang Memproduksi Susu

Kode Emiten	Nama Emiten
SHDA	Sari Husada Tbk
ULTJ	Ultra Jaya Milk Tbk

Tabel 4.7. Kode dan Nama Emiten Perusahaan Makanan dan Minuman yang Memproduksi Minyak Goreng

Kode Emiten	Nama Emiten
SMAR	Smart Corporation Tbk
TBLA	Tunas Baru Lampung Tbk

4.2. Variabel yang dicluster

Pada data saham terdapat empat variabel, yaitu : *previous price*, *highest price*, *lowest price* dan *closing price*. Pada penelitian ini hanya menggunakan dua variabel, yaitu : *previous price* (harga sebelum) dan *closing price* (harga penutupan).

4.3. Normalisasi Data

Pada prakteknya, biasanya perlu dilakukan *preprocessing* terlebih dahulu. Akan lebih menguntungkan apabila data yang akan diolah dalam keadaan normal, misalkan berada pada interval [0,1]. Dengan demikian perlu dilakukan normalisasi terlebih dahulu. Untuk normalisasi metode yang digunakan adalah metode *range* sehingga data yang ada memiliki nilai antara 0 dan 1. Rumus metode *range* :

$$X_{baru} = \frac{X_{lama} - X_{min}}{X_{max} - X_{min}}$$

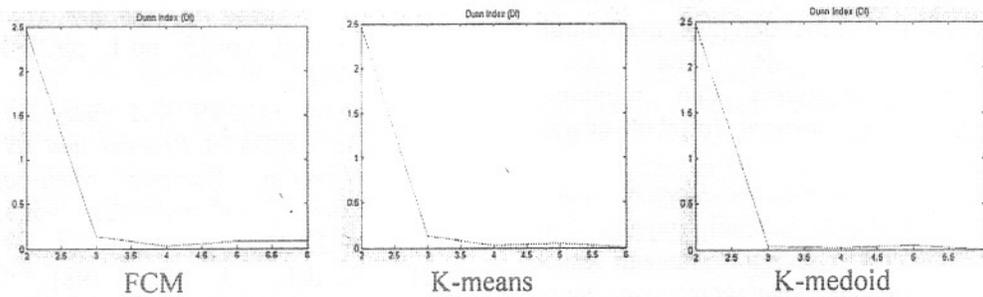
4.4. Hasil Clustering Perusahaan minuman dalam kemasan

Pengklusteran data saham perusahaan makanan dan minuman yang memproduksi air minuman dalam kemasan menggunakan metode *FCM*, *K-Means* dan *K-Medoid*. Hasil clustering dapat dilihat pada tabel 4.8.

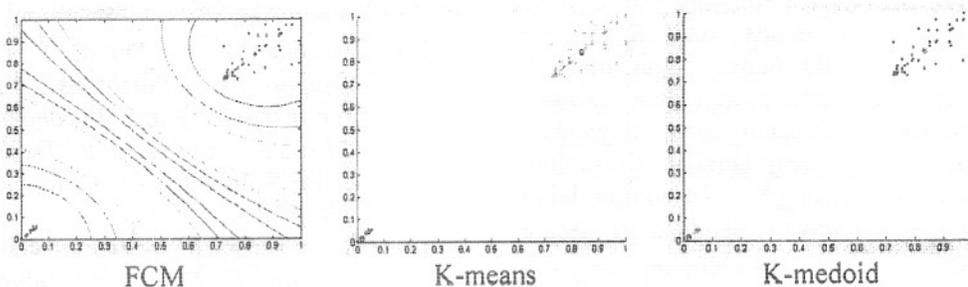
Berdasarkan tabel 4.8 menunjukkan jumlah cluster yang paling optimal untuk ketiga metode tersebut adalah 2 cluster. Metode *FCM* dan *K-means* memiliki iterasi sedikit yaitu sebanyak 2 iterasi, sedangkan metode *K-medoid* memiliki iterasi sebanyak 3 iterasi dengan pusat cluster untuk masing-masing cluster seperti terlampir pada tabel 4.8. Validasi hasil *clustering* untuk tiap-tiap metode dengan visualisasi secara grafik ditampilkan pada gambar 4.1.

Tabel 4.8 : Hasil clustering dengan metode FCM, K-means dan K-medoid

C	FCM				K-means				K-medoid			
	I	V*		DI	I	V*		DI	I	V*		DI
2	2	0.8382	0.8379	2.4643	2	0.0170	0.0171	2.4643	3	0.8346	0.8346	2.4643
		0.0171	0.0171			0.8381	0.8378			0.0171	0.0171	
3	6	0.9275	0.9278	0.1307	4	0.0066	0.0066	0.1307	2	0.7568	0.7568	0.0429
		0.0170	0.0170			0.8381	0.8378			0.9232	0.9232	
		0.7573	0.7570			0.0526	0.0527			0.0171	0.0171	
4	6	0.8385	0.8382	0.0265	4	0.0006	0.0006	0.0265	5	0.0006	0.0006	0.0265
		0.0181	0.0181			0.8381	0.8378			0.8381	0.8378	
		0.0539	0.0539			0.0178	0.0178			0.0178	0.0178	
		0.0007	0.0007			0.0528	0.0529			0.0528	0.0529	
5	8	0.9383	0.9389	0.0794	4	0.0009	0.0009	0.0485	6	0.7568	0.7568	0.0485
		0.0058	0.0058			0.8381	0.8378			0.9232	0.9232	
		0.0523	0.0524			0.0178	0.0178			0.0176	0.0176	
		0.8744	0.8744			0.0002	0.0002			0.0006	0.0006	
		0.7536	0.7534			0.0528	0.0529			0.0507	0.0507	
6	6	0.9944	0.9956	0.0794	7	0.0003	0.0003	2.5254	4	0.7568	0.7568	0.0023
		0.0058	0.0058			0.8381	0.8378	e-004		0.9232	0.9232	
		0.9262	0.9265			0.0200	0.0200			0.0176	0.0176	
		0.0523	0.0524			0.0009	0.0009			0.0010	0.0010	
		0.7530	0.7528			0.0528	0.0529			0.0507	0.0507	
		0.8703	0.8705			0.0169	0.0168			0.0003	0.0003	



Gambar 4.1. Validasi hasil clustering dengan DI



Gambar 4.2. Hasil clustering dengan metode FCM, K-Means dan K-Medoid

Berdasarkan gambar 4.2. dapat diketahui data saham perusahaan makanan dan minuman yang memproduksi minuman dalam kemasan dikelompokkan dengan dua cluster. Cluster pertama merupakan kelompok data yang memiliki harga saham yang tinggi pada variabel *previous price*

dan *closing price* dan cluster kedua adalah kelompok data yang memiliki harga saham yang rendah pada kedua variabel tersebut.

Hasil clustering dengan ketiga metode tersebut menunjukkan yang termasuk dalam cluster pertama adalah Aqua Golden Mississippi Tbk, sedangkan

Ades Alfindo Tbk dan Suba Indah Tbk termasuk dalam cluster kedua. Gambar clustering dengan metode FCM memperlihatkan bahwa titik-titik data yang menjadi anggota cluster tertentu memiliki derajat keanggotaan 1. Plot hasil clustering untuk ketiga metode tersebut dapat dilihat pada gambar 4.2.

5. Kesimpulan

Kesimpulan yang dapat diambil dari hasil pengklusteran data saham perusahaan makanan dan minuman dengan metode *Fuzzy C-Means*, *K-Means* dan *K-Medoid*, maka secara garis besar dapat diperoleh kesimpulan sebagai berikut :

1. Metode *Dunn's Index* dapat digunakan pada ketiga metode *clustering* (*FCM*, *K-Means* dan *K-Medoid*) untuk mengetahui jumlah cluster yang paling optimal.
2. Berdasarkan hasil dari perhitungan *Dunn's Index* ternyata jumlah cluster yang paling optimal dan digunakan untuk pengklusteran pada data saham perusahaan makanan dan minuman dengan ketiga metode tersebut adalah sama.
3. Dengan metode *Fuzzy C-Means*, setiap titik data dapat menjadi anggota dari beberapa cluster secara bersamaan sesuai dengan derajat keanggotaannya yaitu memiliki interval antara 0 sampai 1. Sedangkan dengan metode *hard clustering* (*K-Means* dan *K-Medoid*) setiap titik data hanya dapat menjadi anggota dari satu cluster saja. Dengan demikian pendekatan *crisp* digunakan untuk hal-hal yang bersifat diskontinu. Untuk hal-hal yang bersifat kontinu lebih baik menggunakan pendekatan secara *fuzzy*.
4. Harga saham pada perusahaan makanan dan minuman yang memproduksi coklat bubuk mengelompok secara merata pada setiap cluster, artinya setiap data yang dimiliki oleh masing-masing perusahaan tidak ada perbedaan yang signifikan. Sedangkan untuk perusahaan pada kategori yang lain harga saham antara

perusahaan yang satu dengan yang lain sangat besar perbedaannya.

5. Analisis kluster dilakukan untuk mengelompokkan objek-objek berdasarkan karakteristik yang dimiliki sehingga dapat dikelompokkan. Dalam bidang pemasaran kegiatan tersebut disebut dengan segmentasi (pengelompokan) nilai saham di pasar. Investor dapat memanfaatkan hasil clustering pada perusahaan makanan dan minuman yang terbagi menjadi 7 kategori berdasarkan produknya dalam pengambilan

6. Ucapan Terima Kasih

Terima kasih kepada Prof. Drs. H. Subanar, Ph.D., selaku pembimbing utama dalam penelitian ini.

Daftar Pustaka

- [1] Ruspini, E.H., 1969, *A New Approach to Clustering, Information and Control*, vol.15, no.1, pp.22-32, July 1969.
- [2] Dunn, J.C., 1973, *A Fuzzy Relative of the ISODATA Process and Its Use in Detecting Compact Well-Separated Cluster*, *J.Cybernetics*, vol.3, no.3, pp.32-57
- [3] Bezdek, J. C. 1981, *Pattern Recognition with Fuzzy Objective Function Algorithms*, Plenum Press, New York
- [4] Bezdek, J. C, Dunn, J. C, 1975, *Optimal Fuzzy Partitions : A Heuristic for Estimating the Parameters in A Mixture of Normal Distributions*. *IEEE Transacions on Computers* pages 835-838.
- [5] Hoppner, F., 2003, *Knowledge Discovery from Sequential Data, Identification*, University of Veszprem, Department of Process Engineering, <http://www.fmt.vein.hu/softcomp>
- [6] Bezdek, J. C, Pal, N. R, 1998, *Some New Indexes of Cluster Validity*, *IEEE Trans. Syst. Man Cybern*, B 28, 301-315