Fuzzy c-shells for clustering of data’s LQ45 in Indonesia based on earnings per share and price earning ratio

Nafisa Berliana Indah Pratiwi1, D R S Saputro3, P Widyaningsih3
1-3) Department of Mathematics, Universitas Sebelas Maret, Indonesia.

Corresponding email: nafisaberlianaindahpratiwi@gmail.com

Abstract. Stocks are financial assets which is procurement valuable document of an enterprise that exhibit the ownership. Bursa Efek Indonesia (BEI) has stock index mentioned as Liquid 45 (LQ45). LQ45 includes 45 enterprises with high fundamentals. Stock investors shall use the finance report as primary decision in stocks investment. Profitability ratios is finance ratio component to evaluate the sale, assets, and stocks value. Earnings Per Share (EPS) measures investor perspective in form of returns calculation that will be obtained in certain period. Price earning ratio (PER) describes amount have to be paid by the investor for the profit of the enterprise. Stocks clustering is a necessary decision to gain maximal profits. Stocks clustering determined by clustering algorithm and stocks budge unpredictably then solved by fuzzy concept. The clustering for LQ45 solved by fuzzy c-shells clustering (FCS). FCS clustering for LQ45 results three clusters. First cluster has 7 enterprises with radius of cluster is 1.77, second cluster has 27 enterprises with radius of cluster is 0.37, and the third cluster has 11 enterprises with radius of cluster is 1.55. Xie Beni validity index value 0.4269798. That most minimum Xie Beni value results optimum and valid clusters.

1. Introduction
Clustering is an unsupervised learning since data used are not labelled. Clustering technique can be used in feature space [1]. Clustering is a process grouping data based on information that was obtained from those data then will be utilized to describe an object [2]. Clustering algorithm aims to separate datasets therefore data on similar clusters have maximum likeness and data between another cluster have minimum likeness. That intentions affect the development of previous clustering algorithm with reducing the complexity of multivariate data based on reduction technique for certain clusters with members of each clusters have likeness with one and another [3]. Clustering algorithm consists of hierarchical and non-hierarchical clustering. Hierarchical clustering will form tree structure called as dendrogram by using heuristic separation or merging technique [4]. Non-hierarchical clustering or partitional clustering separate data to specific cluster. Cluster partition divides n objects into c clusters [5]. Non-hierarchical clustering is classified as the hard clustering and soft clustering [6]. Hard clustering or mentioned as crisp clustering is not suitable for data on clusters which has no countable boundary between clusters [5]. Soft clustering enable data to placed in a cluster or more than one cluster. Soft clustering that uses the fuzzy set for clustering process will place data on cluster based on membership value [7].

Fuzzy clustering is the development of soft clustering based on fuzzy sets. Fuzzy clustering utilizes the data’s membership value that will be measured in interval [0,1] [8]. Mostly used fuzzy clustering algorithm is fuzzy c-means algorithm. Fuzzy c-means algorithm is the expansion of hard c-means
clustering [9]. Formed clusters based on distance between many data points and each center of clusters [10]. Fuzzy c-means clustering (FCM) then developed into Fuzzy c-shells clustering (FCS) with additional parameters, the radius of clusters. FCS algorithm uses the centers and the radius of clusters in its processes. FCS clustering algorithm is an algorithm with cluster circles and ellipses prototype.

Stocks are indirect financial asset which is the procure of valuable document from enterprises that declare the belongings of the enterprises. Bursa Efek Indonesia (BEI) has stock index mentioned as Liquid 45 (LQ45). LQ45 is containing 45 enterprises with high fundamental. Investor shall use the financial statements as the base of decision making to invest some stocks. Profitability ratios is a financial ratios component that used to evaluate assets selling, fund and stocks value. Earnings per share (EPS) utilized to measure the investor’s perspective related to amount of returns will be gained in certain period. Price earning ratio (PER) describes the amount have to be paid by the stockholder for every enterprise’s income.

Clustering on stocks is a necessary action for investor to gain maximum profit on certain stocks. Stocks clustering possibly determined by clustering algorithm and stocks budge unpredictably. Those problems can be solved by fuzzy concept. A research on data’s LQ45 using fuzzy clustering was conducted by [11]. Reference [11] used the Fuzzy c-Means and Gustafson Kessel algorithm and both result three clusters which contain the LQ45’s members. Fuzzy c-Means is evolving into Fuzzy c-Shells algorithm then FCS clustering algorithm is also suitable to solve the clustering on data’s LQ45. Therefore, this research aims to determine number of clusters on data’s LQ45 with FCS clustering algorithm.

2. Research Method

This research is conducted by clustering the data’s LQ45 based on EPS and PER indicator using fuzzy c-shells (FCS) clustering algorithm. Data’s LQ45 based on EPS and PER 2018 that were obtained from Indonesia Stock Exchange (IDX). The procedures are shown below.

(1) Describes the descriptive statistics of EPS and PER of data’s LQ45.
(2) Standardize the value of EPS and PER of data’s LQ45.
(3) Determine the clusters using the fuzzy c-shells clustering algorithm (FCS) on data’s LQ45 that have standardized.
(4) Calculates the cluster validity index.

3. Results and Discussion

3.1. Stocks

Stocks are the warrant of fund inclusion in ownership of enterprises. Stockholders on certain stocks divided by single investor (privately owned), small groups (closely owned), or intuition (publicly owned) [12]. Stockholders have rights of the income and wealth of the enterprises after the obligation of the enterprises have paid off.

Stocks index determined based on the fundamental of the enterprise, stocks price or single market capitalization [13]. The index movements describe the whole market movement too. Stock index is an important component as the investment reference to investors. Stock index also helps the investors to determine some actions for example, to sell, hold or buy one or more stocks.

3.1.1. Liquid 45. Index LQ45 is including the 45 enterprises in Bursa Efek Indonesia (BEI) with high liquidity and large market capitalization also fulfill the qualifications according to several selection criteria [14]. The selection criteria of LQ45 are mentioned

(1) Stocks are members of top 60 ranking from stock total transaction in regular market (can determined based on average transaction value in current 12 months)
(2) Stocks are included in top ranks in ranking based on regular market capitalization (can
determined based on average transaction value in current 12 months)
(3) Stocks are registered in Bursa Efek Indonesia (BEI) for at least 3 months
(4) Financial condition of the enterprise and the growth prospect from the owner of the stocks must
be well condition, frequency and the total of transaction days in regular market also in well
condition.

3.2. Financial Ratio
Financial ratio is part of financial statement. Financial ratio is basis of scoring the valuable document
and evaluate the performance of enterprises. The information which are obtained is the wellness of
financial ratio in helping the user to scoring the financial condition of the enterprise and the management
activeness in obtains returns from the fund that have been invested. Financial ratio is grouped as liquidity
ratios, activity ratios, debt ratios, and profitability ratios. Profit is used to measure the enterprise’s
performance based on profits were obtained from an enterprise will affects the enhancement the
investor’s welfare [15]. Enterprise’s profitability analysis can be an important aspect that can be
considered by the investors [16]. Profitability ratios is related to important information for operation
activity of an enterprise in short-term. Profitability ratios is ratio or a comparison to discover the
enterprise’s ability to gain profits from the earnings that related to selling, assets and stocks value based
on certain measurement. Two components of profitability ratios are earnings per share (EPS) and price
earning ratio (PER).

3.2.1. Earnings Per Share. Earnings per share (EPS) is important accounting indicator related to risks,
entity performance and the success of an enterprise [17]. Earnings per share state the number of
incomes which is obtained for every sheets of common stocks that are published.

\[ EPS = \frac{Earnings\ available\ for\ common\ stockholders}{Numbers\ of\ shares\ of\ common\ stock\ outstanding}. \]

An enterprise with high EPS value will affects the enhancement of stocks market. Higher stocks price
will create positive impressions from enterprise’s product on customer side results bigger demand,
higher selling and incomes [18].

3.2.2. Price Earning Ratio. Price earning ratio (PER) is a ratio to declare the value for the enterprises
to measure stocks relative price with the income for every stock [19]. High value of PER can be the
reference for the investors related to the high income that will be gained by the enterprises on the next
year [20].

\[ PER = \frac{Market\ price\ per\ share\ of\ common\ stock}{Earnings\ per\ share}. \]

Higher value of PER, the bigger investor’s reliance to enterprise’s condition in the future.

3.3. Fuzzy c-shells clustering algorithm
Fuzzy c-shells (FCS) clustering algorithm is the development of fuzzy c-means (FCM) clustering
algorithm. FCS and FCM possibly to have similar centre of clusters (\(v_i\)) for some cases [21]. The
difference of FCS and FCM algorithm is FCS adds new parameter (radius) to the algorithm. FCS
algorithm is the first algorithm that uses the non-linear prototype concept based on FCM algorithm [22].
The cluster prototype on FCS algorithm have ellipses and circle shapes that characterized by radius and
centre of clusters. FCS algorithm does not notice the noise on data in clustering process. FCS is fuzzy
clustering algorithm with objective function principle. The purpose of fuzzy clustering based on
objective function is to minimize the objective function, weighted distance from data to clusters [23].
Given \( \mathbb{R} \) is the set of real numbers dan \( \mathbb{R}^p \) is sets of \( p \)-tuples results basic form hyper-spherical shells on center \( v \), \( v \in \mathbb{R}^p \) and radius \( r \). Let \( R \) is \( c \)-tuple and \( R = \{ r_1, r_2, \ldots, r_n \} \in \mathbb{R}^+ \) obtained is a set of basic form of hyper-spherical shells written as
\[
\{ x \in \mathbb{R}^p \mid (x - v)^T I (x - v) = r^2 \}
\]
where \( I \) is a \( p \times p \) identity matrix. Distance between point \( x_k \in \mathbb{R}^p \) with hyper-spherical shells prototype written as
\[
(D_{ik})^2 = (\| x_k - v_i \| - r_i)^2
\]
where
\[
r_i = \frac{\sum_{k=1}^{n} (u_{ik}^m \| x_k - v_i \|)}{\sum_{k=1}^{n} (u_{ik})^m}
\]
and
\[
v_i = \frac{\sum_{k=1}^{n} (u_{ik})^m x_k}{\sum_{k=1}^{n} (u_{ik})^m}.
\]

Let \( U \in M_{fc} \) is fuzzy \( c \)-partition of \( X \) (\( M_{fc} \) is fuzzy \( c \)-partition set of \( X \)) and results function on fuzzy c-shells
\[
J : M_{fc} \times \mathbb{R}^{cp} \times \mathbb{R}^c \to \mathbb{R}^+
\]

Therefore, the objective function on FCS is written
\[
j(U, V, R) = \sum_{i=1}^{c} \sum_{k=1}^{n} u_{ik}^m (D_{ik})^2
\]

Where \( m \) is a fuzziness index with \( m \in [1, \infty) \), \( u_{ik} \) is the membership degree and \( d^2(v_i, x_k) \) is norm of the Euclidean distance.

Fuzzy c-shells algorithm is written as follows:
1. Define numbers of clusters, \( 2 \leq c < n \), with \( n \) is numbers of data points.
2. Define the exponent (weight) = \( m \), \( 1 < m < \infty \).
3. Set the iteration counter \( j = 0 \), Initialize the fuzzy \( c \)-partition \( U^0 \), where \( U \) is a matrix
\[
U = \begin{pmatrix}
\mu_{11}(x_1) & \ldots & \mu_{1n}(x_n) \\
\vdots & \ddots & \vdots \\
\mu_{c1}(x_1) & \ldots & \mu_{cn}(x_n)
\end{pmatrix}
\]

Where initial matrix partition is chosen randomly with \( \mu \) is value based on row and a column, also applicable to \( x \)
4. Calculate the centre of cluster \( (v_i) \) and the radius of cluster \( (r_i) \) with equation below
\[
v_i = \frac{\sum_{k=1}^{n} (u_{ik})^m x_k}{\sum_{k=1}^{n} (u_{ik})^m}, r_i = \frac{\sum_{k=1}^{n} (u_{ik})^m \| x_k - v_i \|}{\sum_{k=1}^{n} (u_{ik})^m}.
\]
5. Calculate distance \( (D_{ik})^2 \). \( D_{ik} \) is distance between \( k \)-feature vector, \( x_k \) and the \( i \)-prototype defined as
\[
(D_{ik})^2 = (\| x_k - v_i \| - r_i)^2
\]
Where $\| \|$ is Euclidean distance, centre $v_t$ and radius $r_t$ from basic form shell cluster.

Update membership iteration at $j^{th}$, $U$ by the equation

$$u_{ik} = \frac{1}{\sum_{j=1}^{n} \left( \frac{D_{ik}^2}{D_{jk}^2} \right)^{1/(m-1)}}$$

(6) Comparing the $U^j$ and $U^{j-1}$. If $|U^j - U^{j-1}| < \varepsilon$, then the process stops. Otherwise $|U^j - U^{j-1}| > \varepsilon$, set $(j = j + 1)$ and turn into step 3.

Based on Pal [24] the best value of $m$ that can be chosen is on interval [1.5-2.5]. Based on the interval, results average and middle points that equals to 2. The value of $m = 2$ is generally picked for fuzzy clustering algorithm. When $m \to 1$ will result the crisp cluster. The lacks of crisp/hard cluster is can not detect the position between the data on centre of cluster or on the boundaries of the cluster [25].

3.4. Xie-Beni Validity Index

Determination the numbers of clusters in fuzzy clustering algorithm is randomly chosen under the constraint $2 \leq c \leq n$. To obtains the optimum numbers of cluster is called as cluster validity. Optimum cluster results well data description, every cluster are separated and compact [26]. Xie-Beni validity index is one of many validity indexes that measures the optimum numbers of clusters. Xie-Beni validity index is validity index with high accuracy in determining the optimum clusters [27]. The index is written as follows

$$I_{XB}(c) = \frac{\sum_{k=1}^{n} \sum_{l=1}^{c} u_{ik}^m d_{ik}^2}{nm \min_{ij} |v_i - v_j|}$$

3.5. Implementation

LQ45 is including the 45 enterprises with high liquidity and large market capitalization. BEI evaluates the stocks performance that included in LQ45 two periods in a year, first period is on February-July and the second period is August-January so then the lists of LQ45 members possibly changes caused by the performance of those stocks. Investors that might become the stockholders have to analyse the financial statement to discover the earnings per share (EPS) and price earning ratio (PER) that affects the stocks price. This research EPS and PER are inspected variables. Scatter plot of data’s LQ45 based on EPS and PER is shown on figure 1.

![Figure 1. Scatter Plot of EPS and PER data’s LQ45](image_url)

The descriptive statistics of data’s LQ45 can be concluded and shown in Table 1.
Table 1. Descriptive statistics of EPS and PER data’s LQ45

|       | Minimum | Maximum | Mean   | Standard Deviation |
|-------|---------|---------|--------|--------------------|
| EPS   | -308.47 | 4049.62 | 533.6047 | 863.8451           |
| PER   | -16.37  | 59.27   | 16.28978 | 12.87957           |

Table 1 results that the value between EPS and PER indicator have large gap in the unit value then decide to do the data standardization for forward calculation. The data standardization aims to transform data into similar scale.

Fuzzy c-shells clustering on LQ45 stocks index is started with cluster $n$ objects, with $n$ is 45 enterprises on $c$ clusters based on EPS and PER indicator. From the research methods, first step based on the algorithm is to determine the numbers of clusters. This research uses the numbers of clusters ($c$), $c = 2, 3, \ldots, 10$. The results are written in Table 2.

Table 2. Numbers of clusters, objective function, and Xie-Beni validity index with FCS algorithm

| Numbers of clusters | Iteration | Objective function | Xie-Beni validity index |
|---------------------|-----------|--------------------|------------------------|
| 2                   | 37        | 168.7656735        | 0.6077625              |
| 3                   | 10        | 30.16173683        | 0.4269798              |
| 4                   | 24        | 99.39281048        | 0.4450277              |
| 5                   | 21        | 92.05076945        | 0.5752448              |
| 6                   | 46        | 141.5233266        | 1.796706               |
| 7                   | 20        | 83.05112915        | 2.182531               |
| 8                   | 100       | 213.6830163        | 1.984883               |
| 9                   | 31        | 89.19481549        | 1.369098               |
| 10                  | 33        | 87.28615965        | 1.328566               |

The simulation of numbers clusters from $c = 2, 3, \ldots, 10$ results the most minimum Xie-Beni validity index 0.4269798 with the numbers of clusters , $c = 3$. Hence, optimum clusters on FCS use in data’s LQ45 are three clusters. Formed clusters shown in Figure 2.

![Figure 2. Optimum clusters based on FCS clustering algorithm on data’s LQ45](image-url)
Three clusters contain 45 enterprises, with first cluster contains 7 enterprises, second cluster contains 27 enterprises and the third cluster contains 11 enterprise. The members of each clusters are written on Table 3.

| Cluster | Members of clusters |
|---------|---------------------|
| 1       | Bank Central Asia Tbk., Gudang Garam Tbk., Indah Kiat Pulp & Paper Tbk., Indo Tambangraya Megah Tbk., Pabrik Kertas Tjiwi Kimia Tbk., United Tractors Tbk., Unilever Indonesia Tbk. |
| 2       | Adaro Energy Tbk., AKR Corporindo Tbk., Astra International Tbk., Bank Negara Indonesia (Persero) Tbk., Bank Rakyat Indonesia (Persero) Tbk., Bank Tabungan Negara (Persero) Tbk., Bank Mandiri (Persero) Tbk., Bank Bumiputra Damai Tbk., Bank Tabungan Pensiunan Nasional Syariah Tbk., Ciputra Development Tbk., Erajaya Swasembada Tbk., XL Axiata Tbk., Indofood Sukses Makmur Tbk., Indika Energy Tbk., Japfa Comfeed Indonesia Tbk. Jasa Marga (Persero) Tbk., Matahari Department Store Tbk., Medco Energi Internasional Tbk., Media Nusantara Citra Tbk., Perusahaan Gas Negara Tbk., Bukit Asam Tbk., PP (Persero) Tbk., Pakuwon Jati Tbk., Surya Citra Media Tbk., Sri Rejeki Isman Tbk., Wijaya Karya (Persero) Tbk., Waskita Karya (Persero) Tbk. |
| 3       | Aneka Tambang Tbk., Barito Pacific Tbk., Charoen Pokphand Indonesia Tbk., H.M. Sampoerna Tbk., Indofood CBP Sukses Makmur Tbk., Vale Indonesia Tbk., Indocement Tunggal Prakarsa Tbk., Kalbe Farma Tbk., Semen Indonesia (Persero) Tbk., Telekomunikasi Indonesia (Persero) Tbk., Chandra Asri Petrochemical Tbk. |

The characteristics of each clusters, radius and center of clusters are shown in Table 4.

| Numbers of clusters | Center of clusters | Radius cluster |
|---------------------|-------------------|----------------|
| Cluster I            | 2.1154317 (EPS)   | 0.6727109 (PER)| 1.7725582      |
| Cluster II           | -0.3063413 (EPS)  | -0.5172991 (PER)| 0.3736076      |
| Cluster III          | 0.8730768 (EPS)   | 1.7106247 (PER)| 1.5525239      |

4. Conclusions
In this paper, the FCS algorithm which used to cluster the data’s LQ45 based on EPS and PER ratio results three optimum clusters. First cluster has members 7 enterprises, second cluster has members 27 enterprises, and the third cluster has members 11 enterprises, with most minimum value of Xie-Beni validity index 0.4269798. Members of first cluster are the enterprises with high EPS and PER value.

Acknowledgements
We would like to acknowledge Universitas Sebelas Maret which provide authors chance to present this paper, Dr. Dewi Retno Sari Saputro as the leader of the mathematical soft computing research team and her expert advices to this research and thanks a lot to all committee KNPMP V 2020.

References
[1] Dave R N 1992 Generalized Fuzzy C-Shells Clustering and Detection of Circular and Elliptical Boundaries. *Journal Pergamon Pattern Recognition* 25(7) 713-21
[2] Tan P N, Steinbach M and Kumar V 2006 Introduction to Data Mining (Boston:Pearson Education)
[3] Bezdek, J C, Ruspini H E, and Keller M E Fuzzy Clustering : A Historical Prespective. *IEE Computational Intelligence Magazine*. 2019
8

[4] Omran, Mahamed G H, Engelbrecht P A and Ayed S An Overview of Clustering Methods. *Intelligent Data Analysis* **11** (2007) 583-605

[5] Jafar, Mohamed O A and Sivakumar R 2013 A Comparative Study of Hard and Fuzzy Data Clustering Algorithms With Cluster Validity Indices. *ERCICA*

[6] Grover N 2014 A study of various Fuzzy Clustering Algorithms. *International Journal of Engineering Research* **3**(3) 177-81

[7] Bora J D, Gupta and Anil Kumar 2014 A Comparative Study Between Fuzzy Clustering Algorithm and Hard Clustering Algorithm *IJCIT*

[8] Abonyi, Benjamin and Mark 2007 Advance in Fuzzy Clustering and Its Application (USA : John Wiley and Sons, Inc., Publication)

[9] J C Bezdek Pattern Recognition with Fuzzy Objective Function AlgorithmsNew York: Plenum Press 1981

[10] Gosh S and D K Sanjay Comparative Analysis of K-Means and Fuzzy CMeans Algorithms *International Journal of Advanced Computer Science and Applications* **4**(4), (2013), 35-39

[11] Lailly R, Suparti and Diah S Analisis Kelompok Dengan Algoritma Fuzzy C-Means dan Gustafson Kessel Clustering Pada Indeks LQ45. Jurnal Guassian **4**(3), (2015), 43-52

[12] Gitman J L Principles of Managerial Finance Tenth Edition Prentice Hall, 2002

[13] Wood P C Non-Capitalization Weighted Stock Market Index And Index Fund Or Funds *Patent Application Publication*, 2006

[14] Tandelilin E Portofolio dan Investasi Yogyakarta:Kanisius, 2010

[15] Islam R M, Khan R T, Choudhury, T T, Adnan, M A How Earning Per Share (EPS) Affects on Share Price and Firm Value, *European Journal of Business and Management* **6**(17), 2014

[16] Nicholson S F Price-Earnings Ratios *Financial Analysts Journal*, **16**(4), (1960), 43-45

[17] Kruse R, Christian D and Marie J L 2007 Advances in Fuzzy Clustering and its Applications (England : A John Wiley and Sons, Inc., Publication)

[18] Wu K, and Yang, M A cluster validity index for fuzzy clustering. *Pattern Recognition Letters*, 2005, 1275-1299

[19] Duo C, Xue L, and Du-Wu C An Adaptive Cluster Validity Index for the Fuzzy C-Means *International Journal of Computer Science and Network Security*, **7**, 2007, 146-156