

Fuzzy subtractive clustering based prediction model for brand association analysis

Imam Djati Widodo*

Industrial Engineering Department, Universitas Islam Indonesia, Yogyakarta, 55584, Indonesia

Abstract. The brand is one of the crucial elements that determine the success of a product. Consumers in determining the choice of a product will always consider product attributes (such as features, shape, and color), however consumers are also considering the brand. Brand will guide someone to associate a product with specific attributes and qualities. This study was designed to identify the product attributes and predict brand performance with those attributes. A survey was run to obtain the attributes affecting the brand. Subtractive Fuzzy Clustering was used to classify and predict product brand association based aspects of the product under investigation. The result indicates that the five attributes namely shape, ease, image, quality and price can be used to classify and predict the brand. Training step gives best FSC model with radii (r_a) = 0.1. It develops 70 clusters/rules with MSE (Training) is $9.7093e-016$. By using 14 data testing, the model can predict brand very well (close to the target) with MSE is 0.6005 and its' accuracy rate is 71%.

1 Introduction

Marketing efforts include creating strategies, marketing tactics and operational activities has very significant role in product success. One of important aspect of marketing is branding management. Brand is important because it is not just a name for the product, but also an identity to differentiate it from the products resulting from other companies. Some studies showed that brand, one important aspect of marketing, has significantly contributed to the success.

Product that has positive brand association tends to get better probability to be purchased in the future [1] and recommended to other customer [2]. It can also increase customer's loyalty [3]. Brand is a symbol of the company name on a product that consistently delivers the features, benefits and services to its customers. Consumers in determining the choice of a product will always consider product attributes (such as features, shape, and color), however consumers are also pay attention on brand [4] [5]. Brand will guide someone to associate a product by selecting information elements in the product selection [6]. This was confirmed by Page and Herr which states that the design and brand have different influences on preferences and quality assessment [7]. Preference is determined by the design, while the quality is determined by both design and brand.

Brand equity will give special value to customers. It can help customers to interpret, process and store information related to the product. Brand equity is built by brand awareness, perceived quality, brand association and brand loyalty [8]. Brand association is claimed to the

most important by some expert because it can drive to the better brand performance [9] to survive in this dynamic environment, brand association should involve institutional theory [10].

Brand association known can be used by consumers to identify other products from a company [11]. Many studies showed that positive brand association positively affects behavior. There are many different techniques to measure and develop brand association network [12] such as brand concept maps (BCM) [13]. However, it is still needed to develop techniques and methods to evaluate and predict the brand association.

Based on some facts and studies above, this research is intended to develop prediction method of the brand association based on customer perception of product attributes. Fuzzy clustering is implemented to divide the data space into fuzzy clusters, in which each representing a specific part of the system behavior. The fuzzy rules can be found for each cluster. By applying fuzzy systems, uncertainty of the case can be involved in linguistic labels to more adapted results [14, 15].

Fuzzy subtractive clustering (FSC) method is applied to get the prediction. The FSC method is used because it is fast method to estimating clusters of the data as basis to get estimation and it is more consistent than fuzzy C Mean method [16, 17].

FSC has been applied in some field of researches such as three phase induction control [18], Non Linear process modeling [19], Data mining [20] and web design [21].

Hand Phone (HP) or smart phone is chosen as the object of this study because it has a high growth rate. Based on research conducted by the Research on Asia

* Corresponding author: imamdjati@uii.ac.id

(ROA), the average rate of growth of HP in Indonesia from 2006 to 2010 are around 17% per year [22]. In fact, surveying HP Nielsen Company Indonesia stated that the average level of ownership HP increased by about 23% per year from 2006 to 2010 [23]. Sales growth is also expected to remain high for the next years. International Data Corporation (IDC) estimates that mobile phone sales will grow about 10%, equivalent to 48.8 million units in 2012 [24]. Besides, HP has a wide range and variety associated with features and price. Number of series HP-powered GSM in the market is more than 180 series. HP's low end product has the most variants. It is about 42% of the total series in the market.

2 Basic Concept

2.1 Brand

The brand is a name, term, sign, symbol, design or a combination of these things, intended to identify the goods or services and to differentiate them from competitors' products. According to Aaker "A brand is a name or symbol discriminatory (such as logo, seal or packaging) with a view to identify the goods or services of one seller or group specific seller thus distinguishing the goods and services produced by the competitor" [8].

Brand equity is an asset that can provide its own value in the eyes of its customers. Asset that it contains help customers interpret, process and store information related to the product and the brand. Besides giving value for the consumer, it also provides the brand equity value for the company. Aaker defines brand equity by "A set of assets (and liabilities) linked to a brand's name symbol that adds to the value provided by a product or service to a firm and for that firm' customers" [8]. Five categories of assets includes brand loyalty, image quality is captured, associations brand, brand awareness and the asset of other brands (other proprietary asset) such as patents, cap, distribution channels and others, all of which can provide value to the customer and the company. Five categories of assets underlying brand equity is illustrated in figure 1.

Romaniuk and Nenycz-Thiel underlined that the knowledge and experience of a consumer with a certain brand will influence the total amount of associations a consumer related to a brand [3]. Based on sport shoes study, Mühlbacher et al. claimed that cognitive brand research showed that brand association characteristics - such as favorability, number, uniqueness, and consensus - influence brand strength [25]. Moreover, the finding supported the theory of cognitive brand equity. Favorability, number, uniqueness, and consensus can be used as predictor of brand strength.

2.2 Fuzzy Subtractive Clustering

The idea of fuzzy clustering is to divide the data space into fuzzy clusters, each of which represents a particular part of the system behavior. After projecting the cluster to the input space, the antecedent part of the fuzzy rules can be found. In the fuzzy clustering algorithm, the membership

function can be determined according to two possible methods [26].

In the first method, the cluster is projected orthogonal to the axis of the antecedent variables and membership functions attached to these projections. The second method uses multi-dimensional membership functions, the fuzzy cluster is projected to input space. Membership degree of data points are directly calculated in this cluster and is projected by distance from the center of the cluster is projected.

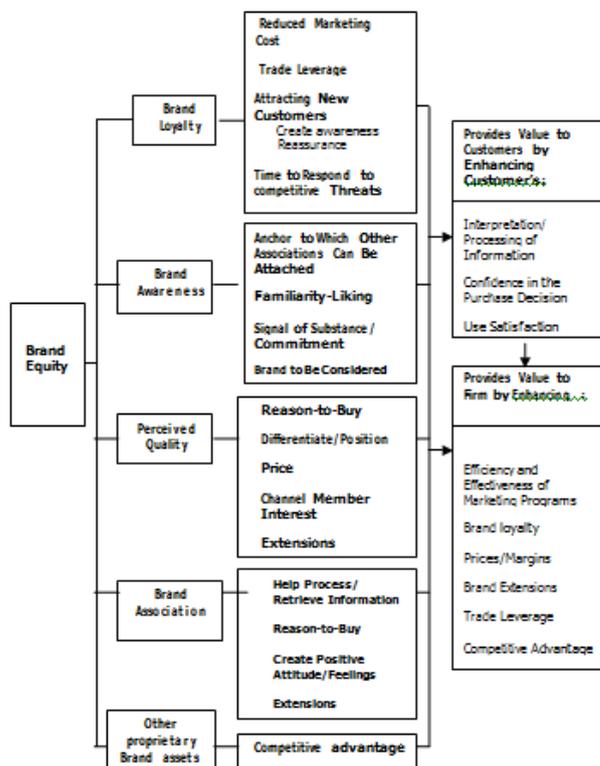


Fig. 1. Brand Equity Value [8]

Some fuzzy clustering methods have been developed in the literature. The most common method is fuzzy C-means clustering and subtractive clustering. Fuzzy C-Means (FCM) is a supervised filtering algorithm. It is developed based on a determined (known) cluster numbers to be performed. If the number of clusters to be formed is not known before, then an unsupervised algorithm should be performed. The most common unsupervised algorithm is Fuzzy Subtractive Clustering (FSC). FSC is based on the size of the density (potential) of data points in a space (variable). Another difference is related to the position of cluster central. In FCM method, the cluster central may not be in one of the clustered data. It will not happen in the subtractive clustering method. In the subtractive clustering method, a cluster center must be one of the clustered data, where the degree of membership in the cluster is equal to 1. The sum of all degrees of membership in FCM is always equal to 1, but not so with subtractive clustering. In the subtractive clustering method, the sum of all degrees of membership is not necessarily (even rarely) equal to 1.

FSC is a fuzzy clustering method that applied Sugeno fuzzy inference System (FIS) and it is unsupervised model. Figure 1 is the structure of the Sugeno FIS.

Subtractive Fuzzy Clustering (FSC), an extension of the mountain clustering method proposed by Yager and Filev is a method that is expected to resolve issues that arise in the previous grouping method in which the calculation is growing exponentially with the dimension of the problem, namely because the function must be evaluated at each grid point.

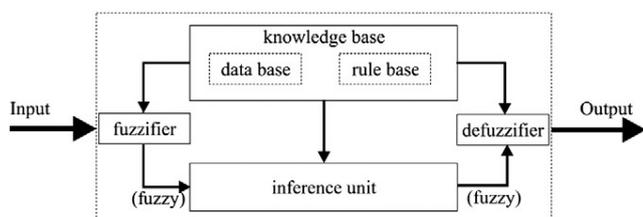


Fig. 2. Structure of FIS

FSC solves the problem by using a data point as a potential center of the cluster, not the grid points as in the mountain cluster. This means that the calculation is now comparable to the dimensions of the problem size; not dimension [27, 28]. Although the exact center of the cluster is not necessarily located at one data point, but in many cases it is a good approach. Each data point is the candidate for the center cluster, the size of the density of data points as shown in (2.1).

$$D_k = \sum_{j=1}^N \exp \left(- \frac{\|X_k - X_j\|}{(r/2)^2} \right), \quad (2.1)$$

when r_a is a positive constant that describes to a radius around a central point.

The first cluster center x_{c1} be selected that have the greatest value of D_{c1} . Then measures the density corrected by equation 2.2.

$$D'_k = D_k - D_{c1} * \exp \left(- \frac{\|X_k - X_{c1}\|}{(r_b/2)^2} \right) \quad (2.2)$$

By D_k is a solid potential for the center cluster to k (c_k). R_b is a positive constant that defines a radius around a central point which has a measurable reduction of solids. R_b value describes the range subtractive efficient and have a greater value than the value r_a .

$$r_b = \eta r_a, \quad (2.3)$$

with η is a value greater than 1 is referred to as squash factor. After subtraction, any potential new cluster centers associated with upper acceptance limit and lower rejection limit and a distance criterion selects the second cluster centers.

In the implementation, there are 2 fractions as a comparator factor. They are Accept ratio and Reject ratio. Accept ratio and reject ratio is a fractional number that has value 0 to 1. Accept ratio is the lower limit where a data point becomes a candidate (candidate) cluster center is allowed to become the center of the cluster. While the reject ratio is the upper limit in which a data point to be a candidate (candidate) cluster center is not allowed to be the center of the cluster. On an iteration, if it has been

found a data point with the highest potential (eg. X_k with potential D_k), then it will be continued by searching for the potential Ratios of the data point with the highest potential of a data point at the beginning of the iteration (eg. X_h with potential D_h). The result for D_k with D_h is then called Ratio (Ratio = D_k / D_h). There are 3 conditions that can occur in an iteration (figure 2):

- If Ratio > Accept ratio, then the data point is accepted as a new cluster center.
- If the Reject Ratio < Ratio ≤ Accept ratio then the new data point will be accepted as the new cluster center only if the data point is located at a considerable distance with the other cluster center (the sum of the Ratio and the nearest distance of the data point with a center other existing clusters ≥ 1). If the sum between the ratio and the longest distance of the data point with the other existing cluster center < 1, then other than that point the data will not be accepted as the center of the cluster, it will no longer be considered to be the center of the new cluster (its potential is set equal to zero).
- If the ratio ≤ Reject ratio, then no more data points will be considered to be a cluster central candidate, iteration is stopped

3 Research Method

This study was designed to cluster the samples and develop FSC model to predict brand of products with regard to the nature of contingency products segment. Research design was explanatory or confirmatory research that aimed to find the relationship of research variable. A survey was conducted to determine the relationship between some HP independent variables and brand choice for low end segment. The independent variables were shape, quality, convenience, price and image.

This research consists of four steps namely data instrument development and instrument testing, data gathering, modeling, and model testing and prediction.

First, research instrument was developed based on some variables as found in the literature review. This study used consumer product attributes preferences questionnaire with Summated Rating Scale Method. A questionnaire was composed to get information from respondents.

Secondly, based on tryout data, instrument was tested for their reliability and validity. Thirty five respondents were involved in tryout. Thirdly, research data gathering was run based on cross-sectional research design type, in which one or more samples were taken from the population at certain of time.

The population of this research was the HP consumers aged between 17-55 years and use low/medium end HP. The sample was taken by purposive sampling method from Yogyakarta and Surakarta. Questions regarding the background of respondents asked in the questionnaire and if it did not meet the criteria will be sorted. Minimum number of samples was determined based on normal assumption (based Kolmogorov Smirnov test). This research used 97 samples. The next steps, FSC modeling

was developed based on data gathered. Eighty-three data was used for model development. Finally, the model was used to predict the target. It was done by some other data (using 14 data). The result was tested to the fact.

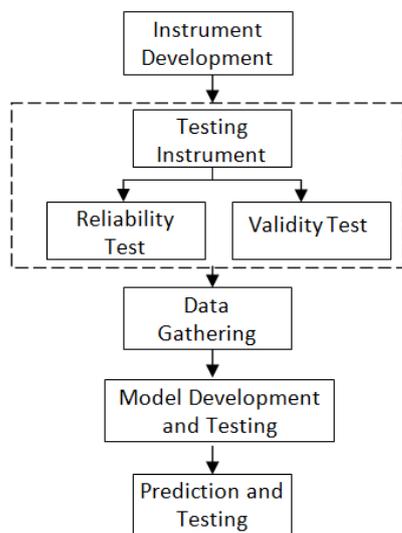


Fig 3. Research flowchart

4 Result and Discussion

There were three steps of analysis in this research; namely validity and reliability analysis, model development, and testing & prediction.

4.1 Validity and Reliability Analysis

Before it was applied, research questionnaire must be analyzed for their validity and reliability. Thirty seven respondents participated in try out step. All items were tested and the result was valid ($r > 0.3$) and reliable ($\alpha=0.8898$).

4.2 Model Development and Evaluation

Data analysis was performed using Matlab 14. The program was done with the help function `genfis2` and `evalfis`. Models were generated by using functions `genfis`. After doing several trials with different radii (r), the best model was found by considering the magnitude of the error rate and number clusters. `Genfis2` was applied to generate a Sugeno-type Fuzzy Interface System (FIS) structure using subtractive clustering for adaptive neuro fuzzy interface system (ANFIS) training. `Evalfis` function was to perform fuzzy interface calculation.

The first step to develop a model is running training. Source code for this model development (training) based on 5 independent variables is performed below:

```

In=P(:,2:6); Out=P(:,7);
fismat=genfis2(In,Out, r);
fuzout=evalfis(In,fismat);
trnmse=norm(fuzout-Out)/sqrt(length(fuzout));
salah=fuzout-Out;
hasil1=[fuzout]
    
```

```

selisih1=[salah]
MSE1=sqrt(mse(salah))
plot(no(:,1),Out,'bo',no(:,1),fuzout,'r*');
pause;
    
```

The study applied Mean Square Error (MSE) to measure the error in this training stage. Best model selection is done by simulating the value of radii. The result can be seen in Table 1.

Table 1. The training evaluation result

Radii (R)	Training	
	# rule	MSE
0.9	2	0.5375
0.7	3	0.5230
0.5	10	0.2522
0.3	51	6.9609e-015
0.1	70	9.7093e-016
0.05	73	5.8138e-016
0.01	73	4.1362e-016
0.001	73	4.1362e-016

Based on the calculation subtractive clustering source code above, the results obtained were very close to the target (brand). In training step, with 84 data, the best model was developed. Fuzzy inference rules were generated by `genfis` function. Table 1 simulates model fuzzy inference rules in training step. Number rules will increase with the decrease of radii. The smaller radii was used, the more clusters will be generated. The fewer clusters were generated, the number rules will be generated and the fewer MSE (training) found. However, rules number was constant for radii less than 0.1. Based on this condition, the study decides to choose model with $ra = 0.1$ for the next prediction. It gives 70 rules/clusters and its MSE is $9.709 \text{ e-}16$.

The prediction and its test source code is performed below.

```

CekIn=Y(:,1:5);
CekOut=Y(:,6);
Fuzcek=evalfis(CekIn,fismat);
plot(CekIn(:,5),CekOut,'bo',CekIn(:,5),Fuzcek,'r*');
pause;
err=Fuzcek-CekOut;
hasil=[Fuzcek]
selisih=[err]
MSE=sqrt(mse(Fuzcek-CekOut))
plot(err,'r*');hold on;plot(err);
pause;
    
```

By using 14 testing data, brand association prediction was done and evaluated based on both prediction MSE and accuracy. The best model (with $ra = 0.1$) gave very good result. The result showed that only four of the fourteen results that was significantly difference (more than 0.5) to the targets. It means the accuracy rate was 71% (Table 2).

Table 2. Brand association prediction accuracy (+/-0.5)

No	Target	Prediction Result	Difference	Similar
1	1	1.6666	0.6666	No
2	2	2.9143	0.9143	No
3	2	2.0000	-0.0000	Yes
4	2	3.3070	1.3070	No
5	2	1.9375	-0.0625	Yes
6	2	3.3501	1.3501	No
7	3	2.7500	-0.2500	Yes
8	3	2.9732	-0.0268	Yes
9	3	2.5877	-0.4123	Yes
10	3	3.0000	-0.0000	Yes
11	3	3.0000	-0.0000	Yes
12	3	2.9906	-0.0094	Yes
13	4	4.0000	-0.0000	Yes
14	4	4.0000	0.0000	Yes

Figure 2 shows how close the prediction and the target visually for radii is 0.1. It looks that prediction is very close to the target. Numerically, it also gives smallest MSE (test) by 0.6005. It can be seen in Table 3.

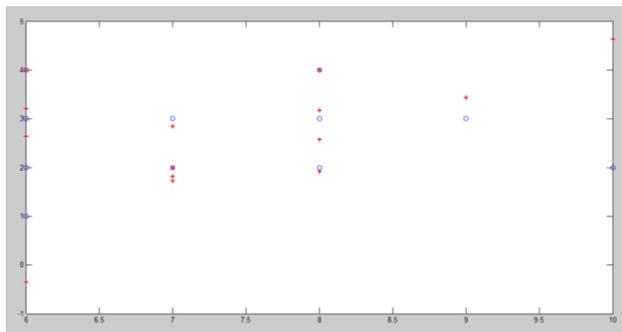


Fig. 2. Comparison between fuzzy subtractive clustering prediction results and target

The prediction performance (MSE Test) and accuracy is influenced much by the characteristic training and testing data.

If the data are similar to the training data, the performance of prediction will be very good. In contrary, if testing data have many outliers, data are higher than the highest training data or lower than the lowest training data, the model built by training data will not predict target of testing data well.

This research also has limitation in determining the best cluster used to get best prediction. It still uses simulation and by trial and error. Thus, further research in optimizing cluster number is needed. Some meta-heuristic such as genetic algorithm, simulated annealing can be applied.

Table 3. The test evaluation result

Radii (R)	# rule	Testing	
		MSE	Accuracy (%)
0.9	2	0.8265	42%
0.7	3	0.8430	42%
0.5	10	2.7409	42%
0.3	51	1.0881	57%
0.1	70	0.6005	71%
0.05	73	0.6467	71%
0.01	73	0.5669	28%
0.001	73	0.5825	28%

5 Conclusion

1. The result indicates that the five attributes namely shape, ease, image, quality and price can be used to classify and predict the brand.
2. Training step gives best FSC model with radii $r = 0.1$. It develops 70 clusters with MSE (Training) is 9.7093e-016.
3. By using 14 testing data, the model can predict brand very well (close to the target) with MSE is 0.6005 and its' accuracy rate is 71%.

References

1. X. Bian, L. Moutinho, The role of brand image, product involvement, and knowledge in explaining consumer purchasing behavior of counterfeits: direct and indirect effects, *European Journal of Marketing*, **45**, 1919-216 (2013)
2. A. Belen, R. Vazquez, V. Iglesias, The effect of brand association on customer response, *Journal of Consumer Marketing*, **18**(5), 410-425 (2001)
3. Romniuk, Nanycz, , Behavioral brand loyalty and consumer brand association, *Journal of Business Research*, **66** (1), 67-72 (2013)
4. Randall, T Ulrich, Andreibstein, , Brand equity and vertical product line extent, *Marketing Science*, **17**(4) 361-365 (1998)
5. G. R. Foxall, V. K. James, The behavioralecology of brand choice: how and what do customers maximize? *Psychology and Marketing*, **20**(9), 811-836, (2003)
6. R. Pieters, L. Warlop, Visual attention during brand choice: the impact of time pressure and task motivation, *International Journal of Research in Marketing*, **16**, 1-16 (1999)
7. C. Page, M. Herr, An investigation of the processes by which product design and brand strength interact to determine initial affect and quality judgments, *Journal Of Consumer Psychology*, **12**(2), 133-147 (2003)
8. A.D. Aaker, Measuring brand equity across products and markets, *California Management Review*, **38**, 3, 102-120(1996)
9. H. He and Y. Li, CSR and service brand: the mediating effect of brand identification and

- moderating effect of service quality, *Journal of Business Ethics*, Vol. 100, 673-688, 2011
10. P. Virutamasena, K. Wongpreedee, W. Kumnungwutab, Strengthen brand association through se: institutional theory, *Procedia - Social and Behavioral Sciences*, 195, 192 – 196 (2015)
 11. D. W. Cravens, *Pemasaran Strategis*, Jilid 2, Edisi 4, Penerbit Erlangga (1998)
 12. S. Gensler, F. Völckner, M. Egger, K. Fischbach, D. Schoder, Listen to your customers: Insights into brand image using online consumer-generated product reviews. *International Journal of Electronic Commerce*, **20**(1), 112–141, (2016)
 13. D. Bögera, P. Kottemanna, M. Meißnerb, R. Deckera, A mechanism for aggregating association network data: An application to brand concept maps, *Journal of Business Research*. 79, 90–106, (2017)
 14. A. Y. Chamzini, M. Razani, S. Yakhchali, E.K. Zavadskas, Z. Turskis, Developing a fuzzy model based on subtractive clustering for road header performance prediction, *Automation in Construction*, 35, 111–120, (2013)
 15. R. Pereira, A. Fagundes, R. Melício, V.M. Mendes, J. Figueiredo, J.C. Quadrado, Fuzzy subtractive clustering technique applied to demand response in a smart grid scope, *Procedia Technology*, 17, 478 – 486 (2015)
 16. K.M. Bataineh, M. Najia, M. A. Saqera, A comparison study between various fuzzy clustering algorithms, *Jordan Journal of Mechanical and Industrial Engineering*, **5**(4), 335 – 343 (2011)
 17. U. M. RAOA, Y.R. Soodb, K. Jarialb, Subtractive clustering fuzzy expert system for engineering applications, *Procedia Computer Science*, 48, 77 – 83 (2015)
 18. J. Dhandapani, R. Chandralekha, Analysis of faults in three phase induction motor using neuro fuzzy logic and wavelet transformation, *International Journal of Control Theory and Applications*, **10**(34), 1-14 (2017)
 19. A. M. Mohan, and M. J. Rani, Application of Soft Computing Techniques for Non-linear Process Modeling, *International Journal Of Innovative Research In Technology*, **2** (10) (2017)
 20. M. Castañon-Puga, J. Flores-Parra, J. Castro, C. Gaxiola-Pacheco, and L. E. Palafox-Maestre, Building java intelligent applications data mining for java type-2 fuzzy inference systems, *Procedia Computer Science*, 51, 2719–2728 (2015)
 21. D. S. Sisodia, S. Verma, O. P. Vyas, A subtractive relational fuzzy c-medoids clustering approach to cluster web user sessions from web server logs, *International Journal of Applied Engineering Research*, **12**(7), 1142-1150 (2017)
 22. Firdaus, *Bisnis ICT di Indonesia, Siapa yang Menikmati?*, *Harian Bisnis Indonesia*, 17 Januari 2012
 23. Nugroho, F, *Perkembangan Pasar Handphone di Indonesia Dari Tahun 2005 Hingga 2010*, available: <http://www.teknojurnal.com/2011/03/03/perkembangan-pasar-handphone-di-indonesia-dari-tahun-2005-hingga-2010/> (2012)
 24. A. Wicaksono, *IDC perkiraan penjualan ponsel tumbuh 10%*, available: <http://industri.kontan.co.id/news/2012-idc-perkiraan-penjualan-ponsel-tumbuh-10> (2012)
 25. H. Mühlbacher, K. Raies, R. Grohs, O. Koll, Drivers of brand strength: Configural paths to strong cognitive brand equity. *Journal of Business Research*, 69, 2774–2780, (2016)
 26. A. Priyono, P. Ridwan, A.J. Alias, R. Atiq, O. K. Rahmad, A. Hassan, and A. M. Ali, Generation of fuzzy rules with subtractive clustering, *Jurnal Teknologi*, 43(D), 143–153 (2005)
 27. R.S. Sidhu, K. Sunil, P. S. Sandu, R. P. S. Bedi, K. Kaur, A subtractive clustering based approach for early prediction of fault proneness in software modules, *Engineering and Technology*, 67, 828-832, (2010)
 28. M. T. Hayajneh, A. M. Hassan, F. Al-Wedyan, Monitoring defects of ceramic tiles using fuzzy subtractive clustering-based system identification method, *Soft Comput*, Mey, Springer, and Online Published (2009)