Proceedings of 8th IOE Graduate Conference
Peer Reviewed
ISSN: 2350-8914 (Online), 2350-8906 (Print)
Year: 2020 Month: June Volume: 8

Social Media as a Strategic Marketing Tool to Study the Consumer Behavior of Coca-Cola by K-Means Cluster Analysis: Evidence from Facebook

Sumit Chaulagain [a], Sanjeev Maharjan [b]

[a,b] *Department of Mechanical Engineering, Pulchowk Campus, IOE, TU, Nepal*
Corresponding Email: [a] sumitchaulagain.2@gmail.com, [b] sanjeevworldnpl@gmail.com

Abstract
Social media is the collective of online communications channels leading to a deep transformation of the organizational models, community-based input, interactions, content sharing collaborations and changing the company's marketing dynamics. This study aims to recognize the consumer behavior of Coca-Cola and validate consistency of respondents. Sets of questionnaire are developed for understanding the consumer behavior and activities in Facebook. The consistency of respondent of consumer analysis were viewed through K-means clustering technique using through SPSS version 16.0 software. Groups of similar and dissimilar of users attitude regarding the product is categorized by the small distances among the cluster members, dense areas of the data space and the intervals of particular statistical distributions. The respondents were categorized into four clusters using two variables: usage rate and preference. Cluster 1 were labeled consumers with high usage rate and highly preference of Coca-Cola, whereas Cluster 2 were labeled consumers with high usage rate and low preference of Coca-Cola. Cluster 3 were labeled consumers with low usage rate and high preference of Coca-Cola and Cluster 4 was labeled with low usage rate and low preference of Coca-Cola. Majority of respondents are categorized under Cluster 3 while minority of respondents are categorized under Cluster 2 by using K-mean cluster. Finding of this research may facilitate the Nepalese business organizations for formulating policies of sales promotion on soft drinks.

Keywords
social media, consumer behavior, consistency, respondents, K-means clustering

1. Introduction

Social media is a phenomenon that has drawn a lot of attention both to companies and individuals interacting on the networking landscape. Social media has influenced consumer behavior from information acquisition to post-purchase behavior such as dissatisfaction statements or behaviors about a product or a company [1]. Social media websites provide an opportunity for businesses to engage and interact with potential consumers, encourage an increased sense of intimacy with consumers, and build all important relationships with potential consumers [2]. The much higher level of efficiency of social media compared to other traditional communication channels prompted industry leaders to state that companies must participate in Facebook, Twitter, Myspace and others, in order to succeed in online environments [3]. Amongst the registered online social network, Facebook is the biggest social network platform in the world with more than 2.49 billion active users, followed by Youtube (about 2 billion active users) and WhatsApp (about 1.92 billion active users). The Facebook platform enables the exchange of information quickly, flexibility and easily [4].

Due to globalization, Nepal is also stepping into the shoes of western countries. It is important to study how people utilize social media in their daily lives. Maintaining interpersonal connectivity between online users of a social media channel has benefits derived from establishing and maintaining contact with other people in a manner of giving social support, friendship, and intimacy [5]. Cultural aspects have an influence on consumer's usage of social networks and a great impact over the online purchase intentions [6]. These allow users to connect with peers by adding them to networks of friends, which facilitates communication, particularly among peer groups [7]. In Nepal, many firms have started using these technologies for promoting their products and services, to communicate with their customers, to gain new business leads and even to do market research. Social media platforms offer many other things than just communication alone. That is the reason why it makes social media so important tool in today‘s environment.

One of the most useful techniques in business analytics for the analysis of consumer behavior and categorization is customer segmentation. By using clustering techniques, customers with similar means, end and behavior are group together into homogeneous cluster [8]. Customer segmentation helps organizations in identifying or revealing distinct groups of customers who thinks and function differently and follow varied approaches in their spending and purchasing habits. So, K-mean cluster is the most appropriate for the business analysis in the context of customer segmentation.

2. Literature Review

Many studies show that firms adoption of marketing strategies based on social media have a positive impact related to direct interaction with consumers; the strategies allow the firms, based on consumer feedback, to acquire marketing information and to learn about current performance and predict future performance [9, 10]. Four channels through which social media affects the firm's performance are the relationship between firms and society (social capital), knowledge of consumer preferences (revealed preferences), transformation of social-marketing resources into financial performance capabilities (social marketing) and conversion of social corporate resources into operational performance capabilities (corporate social networking) [11].

Companies use social platforms, such as Facebook and Twitter in order to increase brand awareness and enlist people's participation through online comments, posts, and other types of engagement [12]. The five commonly used approaches to analyze multiple-choice test data are classic test theory, factor analysis, cluster analysis, item responses theory and model analysis. Amongst these analysis, the cluster analysis is a good method to point out how consumer's response patterns differ so as to classify consumer's behavior [13]. There is a positive correlation between the number of people talking about a company on Facebook and the firm's net revenue and number of personnel, indicating that people are talking more about those companies that are larger and more profitable [14]. Similar results were obtained and social media activity on Facebook by South Korean companies was positively and significantly associated with an increase in firm performance measured by financial returns [15].

A small amount of negative information from a few postings can have substantial impacts on consumer attitudes [16]. It was stated that individuals have an underlying need for an emotional bond with high-involvement products that they buy. Brand development and relationship development are complementary and substituting strategies towards this bonding. The relationship between the brand and the consumer is very important for the company and strengthening it leads to loyalty [17].

3. Research Methodology

The research is based on both qualitative and quantitative methods derived from primary and secondary sources. Understanding the consumer behavior is the major problem of research. Secondary data is collected from firm's official Facebook site and primary data is collected from questionnaire related to the consumers behavior and their engagement in Facebook sites. The set of closed questions are developed and forwarded to the respondents or consumers that follows Coca-Cola official website. Sample of 74 respondents of Nepal are taken for analysis.

K-mean clustering of the respondents are performed based on two variables: usage rate and preference, based on score of 1 (Low) to 5 (High). The respondents are categorized into 4 clusters using SPSS version 16.0 software and their demographic profile, Facebook engagement and consumer's behavior pattern are studied. Chi-square test and Phi Coefficient of the data are calculated in order to check goodness of fit and correlation of datas. Lastly, conclusion were drawn and the management committee are suggested for making the selection decision in improving the product evaluation.

3.1 Cluster Analysis

Cluster analysis is one of the most common exploratory data analysis technique which partitioned given dataset into the subset of similar data points in each subset and dissimilar to data from other clusters [18]. It is based on various kinds of object's differences and uses distance functions' regulations to make model classification [19]. Clustering techniques have become very useful for large datasets even in social media such as facebook and twitter [20]. There are four types of clustering techniques: partitioning methods (K-mean method and K-medoids method), hierarchical methods (agglomerative approach and divisive clustering), density based method (DBSCAN algorithm) and grid based methods (CLIQUE algorithm) [21].

K-means clustering technique is a centroid-based iterative method and is widely used algorithms for clustering [22]. In K-mean algorithm, the 'n' number of observations is divided into 'k' clusters such that the observations in a cluster are nearest to each other in reference value like cluster mean and the distance of the object. When used in conjunction with other algorithms like Lloyd's algorithm, etc., the k-mean methods can be applied to large data sets also [23].
The data in K-means are classified in advance into k clusters to define the k-centroid value of each cluster. The location of centroid is of paramount importance since it may give different results when the farther they are the better it is. In the subsequent steps, the data points that belong to a set are moved towards the nearest centroid so that no point remains unmoved [24]. The new k centroids are recalculated many times over so that the dataset belonging to one cluster may switch into another cluster at the time of new clustering. This process is repeated until no possibility of switching over of dataset remains.

K-means Algorithm

This algorithm is performed in the following steps [25]

Step 1: Choose 'k' numbers of clusters to be determined.

Step 2: Choose centroids randomly as the initial centers of the clusters. Start with a set of cluster centroids: C_1, C_2, C_k

Step 3: Computation of cluster centers

• Assign each object to their closest cluster center using Euclidean distance between the data vector X_i and centroid 'C_j'.

The Euclidean distance between an object and all the nearby centroid is calculated as per the formula [26]

$$d_{ij} = \Sigma_{j=1}^{k}\Sigma_{i=1}^{n}[[X_i^{(j)} - C_j]]^2$$

where $[[X_i(j) - C_j]]^2$ is the nearest distance measure between a data point X_{ij} and the centroid C_j, and it indicates the distance between data points from their centroid.

• Compute the membership grades λik. Here, λ_{ik} indicates the amount of association of data vector X_i with centroid C_j and depends on the distance d_{ij}

• Compute new cluster center by calculating mean points.

$$C_k = \frac{\Sigma\lambda_{ik}X_i}{\Sigma\lambda_{ik}}$$

• If the data point is closest to its own cluster, leave it where it is. If the data point is not closest to its own cluster, move it into the closest cluster.

Step 4: Iterate until membership grades remains unchanged, i.e no change in cluster center or no object change its clusters. The final centroids of the k clusters are then determined.

Step 5: Partition 'n' number of data points into k groups; G = G_1, G_2,.....,G_k by minimizing the function within clusters.

$$Chi - Square(\chi^2) = \Sigma\frac{(f_o - f_e)^2}{f_e}$$

$$Phi - Coefficient(\phi) = \frac{\chi}{\sqrt{n}}$$

where, f_o = observed frequency and f_e = estimated frequency

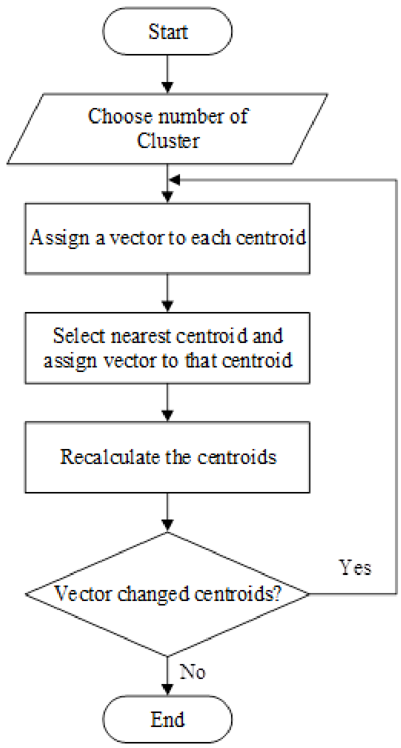

Figure 1: Algorithm of K-Means Clustering

4. Results and Discussion

4.1 Firms Facebook Activities

Coca-Cola has its official Facebook page of ”Coca-Cola Nepal” with 106,683,947 fans or followers. Firms Facebook activities are analysis in the basis of firm’s nature of post (photos and videos) and consumers activities (likes, comments and shares) in Facebook as social media site.

Table 1: Descriptive Data of Facebook Effort for Coca-Cola

	Mean	Min.	Max.	Std. Dev.
Posts/Day	0.296	0.065	0.742	0.187
Likes/Post/Day	118.5	14.726	501.908	134.326
Comments/Post/Day	1.384	0.270	6.452	1.744
Shares/Post/Day	1.558	0.181	6.189	1.633

Table 1 briefly presents data on Firm’s and consumer’s Facebook activity between April, 2019 and March, 2020. In particular, the number of posts per day reached a maximum value of 0.742 with average value of 0.296 while post received an average of 118.5 likes, 1.384 comments and 1.558 shares.

Table 2: Firms Nature of Post

	Mean	Min.	Max.	Std. Dev.
Photos/Day	0.153	0.032	0.323	0.084
Videos/Day	0.129	0.000	0.387	0.123

Table 2 summarizes the nature of the posts published by Firms on Facebook. The firms preferred to use average post of 0.153 photos per day (ranging from 0.032 to 0.323) and 0.129 videos per day (ranging from 0 to 0.387).

4.2 K-Means Clustering

Based on the consumer’s responses in two variables (Usage Rate and Preference), K-mean cluster analysis is performed. These variables are scored from 1 to 5 (1=Low to 5=High) and 4 clusters (k=4) are formed. Table 3 represents four initial centers are chosen for K-mean clustering in SPSS version 16.0 software.

Table 3: Initial Cluster Centers

	Center-1	Center-2	Center-3	Center-4
Usage Rate	5	5	2	1
Preference	5	1	4	1

The final clusters are obtained through iterative process and is shown in Table 4.

Table 4: Final Cluster Centers

	Center-1	Center-2	Center-3	Center-4
Usage Rate	4	4	2	2
Preference	4	2	4	2

Euclidean distance is calculated between the input data and final clusters centers, and 74 respondents are categorized into 4 clusters based on the minimum distance between these centers. Table 5 shows the number of respondents in each clusters. Cluster 3 has majority of respondents (n=35) while cluster 2 has least number of respondents (n=4).

Table 5: Number of Cases in Each Cluster

	Usage Rate	Preference	Frequency (n)
Cluster 1	4	4	25
Cluster 2	4	2	4
Cluster 3	2	4	35
Cluster 4	2	2	10

Consumers with high usage rate and highly preference of Coca-Cola are categoried in Cluster 1. They are loyal consumers of the product and has somewhat stereotype buying behavior (Habitual Buying Behavior).
While, consumers with high usage rate and low preference of Coca-Cola are under Cluster 2. They are negative promoter of the product and have dissonance reducing buying behavior. Improvement of product features like taste, ingredients, variety, etc. are required for booming sales of the product.
In Cluster 3, consumers with low usage rate and high preference of Coca-Cola. They are the consumers that highly prefer Coca-cola but they have low involvement in consumption of soft drinks. Effective marketing with consistent quality is required to improve its involvement.
While, Consumers with low usage rate and low preference of Coca-Cola are categorized in Cluster 4. They do not play any effective roles in sales of the product. These consumers need to be neglected.

Table 6: ANOVA of K-Mean Clustering

	Cluster		Error		F	Sig.
	Mean Sq.	df	Mean Sq.	df		
Usage Rate	27.846	3	0.264	70	105.58	.000
Preference	26.132	3	0.389	70	67.15	.000

Analysis of variance between and within clusters the result of Fishers test (F) and the model significance are shown in Table 6. The result show that the four clusters are significantly different and that all variables are significant at 99% level, as can be seen usage rate is the most influential variable in the profile of cluster (F=105.58) while features rating is the social media dimension with least influence (F= 67.15).

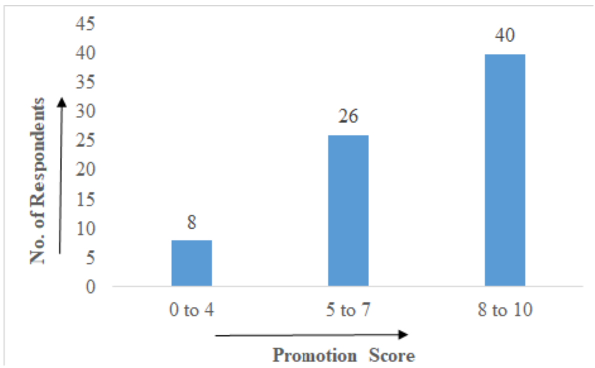

Figure 2: No. of Respondents and Promoting Score

Figure 2 depicts the scoring pattern of respondents for recommending the products to their groups and communities. The promotion score ranges from 0 (Least) to 10 (Most). The detractors, neutral and promoters scores are 0 to 4, 5 to 7 and 8 to 10 respectively. There is large number of promoters (40 respondents) and least number of detractors (8 respondents).

Net Promoter Score (NPS) = Promoters - Detractors

Net Promoter Score (NPS) = 40 - 8 = 32

Table 7: Usage Rate vs Preference of Coca-Cola

Variables	n	C-1	C-2	C-3	C-4
1. Usage rate (χ^2=64.5948 and ϕ=0.9342)					
A. Never	5				5
B. Only for certain occasion	27			22	5
C. Every other week	14	2	1	11	
D. 2-6 times a week	19	17	1	1	
E. Everyday	9	6	2	1	
2. Preference (χ^2=64.5248 and ϕ=0.9338)					
A. Very Poor	6		2		4
B. Poor	9	1	2		6
C. Average	12	1		11	
D. Good	29	8		21	
E. Excellent	18	15		3	

Table 7 shows cluster membership was significantly related with usage rate and preference of Coca-Cola. The strongest association was between usage rate and cluster membership (Chi-Square= 64.5948 and Phi Coefficient = 0.9342) where 27 respondents in the high involvement cluster were using it for certain occasion and 5 respondents in low involvement cluster have never used the product. Relationship between preference and cluster membership shows that 29 respondents find the product good while 6 respondents in low index find the product very poor (Chi-square= 64.5248 and Phi coefficient = 0.9338).

Table 8: Demographic Profile of Respondents

Variables	n	C-1	C-2	C-3	C-4
1. Gender(χ^2=3.9709 and ϕ=0.2316)					
A. Male	46	15	2	20	9
B. Female	28	10	2	15	1
2. Age Group (χ^2=6.0499 and ϕ=0.2859)					
A. 16-20 Years	7	4		3	
B. 20-25 Years	21	6	2	10	3
C. 25-30 Years	23	6	1	12	4
D. 30-35 Years	14	6	1	6	1
E. Above 35 Years	9	3		4	2
3. Education Background (χ^2=67.5630 and ϕ=0.955)					
A. Below SLC/SEE	2	1		1	
B. SLC/SEE	3	1	1	1	
C. +2 or equivalent	21	10		8	3
D. Undergraduate	24	5	3	13	3
E. Postgraduate	24	8		12	4
4. Marital Status (χ^2=7.2720 and ϕ=0.3135)					
A. Single	41	13	2	19	7
B.Married with no children	11	2	2	6	1
C. Married with children	22	10		10	2
5. Occupation or Profession (χ^2=13.4285 and ϕ=0.4260					
A. Student	18	7	1	8	2
B. Employed	39	10		22	7
C. Self-employed	15	7	3	4	1
D. Unemployed	2	1		1	

Table 8 shows significant relationship between cluster membership with all demographic variables. The strongest association was between education background and cluster membership (chi-square = 67.5630 and phi coefficient = 0.955) where 24 respondents each were undergraduate and postgraduate, and 2 respondents were under SLC/SEE. While age group and gender are weakly associated with the cluster membership (chi-square = 6.0499, phi coefficient = 0.2859, and chi-square = 3.9709, phi coefficient = 0.2316 respectively).

Table 9 shows cluster membership was significantly related with respondent's engagement in Facebook. The strongest association was between frequency of usage and cluster membership (chi-square = 13.3124 and phi coefficient = 0.4241) where 36 respondents uses Facebook several times a day while 2 respondents uses Facebook once every couple of months. Number of Facebook friends is the next variable that is strongly associated with cluster membership (chi-square = 9.3855 and phi coefficient = 0.3561). While time spend in Facebook was weakly associated with the cluster membership (chi-square = 2.6855, phi coefficient = 0.1905).

Table 9: Respondent's Engagement in Facebook

Variables	n	C-1	C-2	C-3	C-4
1. Experience in using Facebook (χ^2=7.9049 and ϕ=0.3268)					
A. less than 5 Years	11	6		4	1
B. 5-10 Years	26	6	1	17	2
C. Above 10 Years	37	13	3	14	7
2. Frequency of Usage (χ^2=13.3124 and ϕ=0.4241)					
A. Rarely	3	2			1
B. Once every couple of months	2	2			
C. Atleast once a week	8	4		3	1
D. Several times a week	6	1		4	1
E. Atleast once a day	19	7	2	7	3
F. Several times a day	36	9	2	21	4
3. Time spent in Facebook (χ^2=2.6855 and ϕ=0.1905)					
A. Less than an hour	22	7	1	10	4
B. 1-3 hours	38	13	2	20	3
C. More than 3 hours	14	5	1	5	3
4. No. of Facebook friends (χ^2=9.3855 and ϕ=0.3561)					
A. Less than 200 friends	5			3	2
B. 200-1000 friends	46	16	3	23	4
C. 1000-2000 friends	19	8	1	6	4
D. More than 2000 friends	4	1		3	
5. Purpose of using Facebook (χ^2=4.8468 and ϕ=0.2559					
A. Communication	22	8	1	12	1
B. Entertainment	22	6	1	11	4
C. Information sharing	12	5		5	2
D. All of above	18	6	2	7	3

5. Conclusion and Recommendation

In this paper, we propose an approach to reveal the consumer behavior of Coca-Cola buyers using K-means cluster techniques. For the analysis, 74 respondents are categorized into 4 clusters based on usage rate and preference in SPSS version 16.0 Software.
Cluster 1 includes consumers with high usage rate and high preference. These are loyal consumers and includes 33.78% of respondents. Cluster 2 includes minority of respondents (5.41%), are detractors and have dissonance reducing buying behavior. Cluster 3 includes majority of respondents (47.30%) with less consumption and high preference of Coca-Cola product. Cluster 4 includes consumers with low usage rate and low preference contributes 13.51% of respondents.The consumers of Coca-Cola for each cluster are analyzed and some suggestions of high preference and high usage solutions are provided to prevent them from not purchasing. Net Promotion Score of these respondents is 32 (43.24%). So, majority of people are satisfied with Coca-Cola and promote the product.

This study has both theoretical and managerial implications. From theoretical point of view, this paper helps to bridge the gap related to measure companies' social media's efforts. From a managerial perspective, this study is useful to compare the Coca-Cola's with key competitors. In this sense, it is crucial to focus on the quality of the content of messages and posts, the style of writing and the timeliness of information exchange – all factors that can expand consumer awareness and brand loyalty, with the chance to build deals.

References

[1] G. Mangold and D. Faulds. Social media: The new hybrid element of the promotion mix. *Business Horizons*, 42:357–365, 2009.

[2] R. Mersey, E. Davis, E. Malthouse, and B. Calder. Engagement with media. *Journal of Media Business Studies*, 7(2):39–56, 2010.

[3] A. Kaplan and M. Haenlein. Users of the world unite! the challenges and opportunities of social media. *Business Horizons*, 53:59–68, 2010.

[4] A. Mozas-Moral, E. Bernal-Jurado, M.J. Medina-Viruel, and D. Fernandez-Ucles. Factors for succeess in online social networks: An fsqca approach. *J. Bus. Res.*, 69(11):5261–5264, 2016.

[5] M.D. Utpal, B. Pearob, and L. Klein. A social influence model of consumer participation in network and small-group-based virtual communities. *International Journal of Research in Marketing*, 21:241–263, 2004.

[6] S. Pookulangaran and K. Koesler. Cultural influence on consumer's usage of social networks and its impact on online purchase intentions. *Journal of Retailing and Consumer Services*, 18:348–354, 2011.

[7] P. Ahuja and D. Galvin. Socialization in virtual groups. *Journal of Management*, 29(2):161–185, 2003.

[8] M. Ghanzanfari, S. Malek Mohammadi, and S. Alizadeh. Customer segmentation in clothing exports based on clustering algorithm. *Iranian Business Studies Quarterly*, 14(56):59–86, 2010.

[9] D.G. Gelb and Sundaram. Adapting to word of mouth. *Bus. Horiz*, 14(4):21–25, 2002.

[10] S. Kim, Y. Koh, J. Cha, and S. Lee. Effects of social media on firm value for u.s. restaurant companies. *Int. J. Hosp. Manag.*, 49(1):40–46, 2015.

[11] J. Paniagua and J. Sapena. Business performance and social media: Love or hate? *Bus. Horiz.*, 57(6):719–728, 2014.

[12] E. Wallace, I. Buil, and de Chernatony. Consumer engagement with self-expressive brands:brand love and wom outcomes. *J. Product Brand Manag.*, 23(1):33–42, 2014.

[13] L. Ding and R. Beichner. Approaches to data analysis of multiple-choice questions. *phys. Rev. ST Phys. Educ. Res 5, 020103*, 2009.

[14] H. Karjaluoto, H. Makinen, and J. Jarvinem. A firm's activity in social media and its relationship with corporate reputation and firm performance. *In Procedeeings of the 28th Bled eConference: e WellBeing*, pages 469–481, 2015.

[15] S. Chung, A. Animesh, K. Han, and A. Pinsonneault. Firm's social media efforts, consumer behavior and firm performance. *In Procedeeings of 35th International Conference on Information Systems*, 2002.

[16] A. Schlosser. Posting versus lurking: Communicating in a multiple audience context. *Journal of Consumer Research*, 32(2):260–265, 2005.

[17] A. Palmer. Integrating brand development and relationship marketing. *Journal of Retailing and Consumer Services*, 3:251–257, 1996.

[18] H. Jaiwei, M. Kamber, and J. Pei. *Data Mining: Concepts and Technique*. San Fransciso, California: Morgan Kaufmann Publishers, 2012.

[19] B. Zhaoqi and Z. Xuegong. *Pattern Recognization*. Beijing: Tsinghua University Press, 2000.

[20] M.V. Joseph. Significance of data warehousing and data mining in business applications. *International Journal of Soft Computing and Engineering*, 3(1):719–728, 2013.

[21] A. Fahad, N. Alshatri, Z. Tari, A. Alamri, Y. Zomaya, I. Khalil, S. Foufou, and A. Bouras. A survey of clustering algorithms for big data:taxonomy and empirical analysis. *IEEE Transactions on Emerging Topics in Computing*, 2(3):267–279, 2014.

[22] A. Jain. Data clustering: 50 years beyond k-means. *19th International Conference in Pattern Recognition*, 2009.

[23] R. Agrawal, J. Gehrke, D. Gunopulos, and P. Raghavan. *Automatic Subspace Clustering of High Dimensional Data for Data Mining Applications*. 650 Harry Road, San Jose, CA: IBM Almaden Research Center, 1998.

[24] T. Madhulatha. An overview on clustering methods. *IOSR Journal of Engineering*, 2(4):719–725, 2012.

[25] P.D. Arora. Analysis of k-means and k-medoids algorithm for big data. *International Conference on Information Security and Privacy*, pages 507–511, 2015.

[26] K.S. Kadam and S.B. Bagal. Fuzzy hyperline segment neural network pattern classifier with different distance metric. *International Journal of Computer Applications*, 95(8):6–11, 2014.