TELKOMNIKA, Vol.17, No.3, June 2019, pp.1409~1416
ISSN: 1693-6930, accredited First Grade by Kemenristekdikti, Decree No: 21/E/KPT/2018
DOI: 10.12928/TELKOMNIKA.v17i3.9101

Clustering analysis of learning style on anggana high school student

Siti Lailiyah[1], Ekawati Yulsilviana[2], Reza Andrea*[3]
[1,2]STMIK Widya Cipta Dharma, Samarinda Ulu, Kota Samarinda, Kalimantan Timur, Indonesia
[3]Politeknik Negeri Pertanian Samarinda, Samarinda Utara, Kota Samarinda, Kalimantan Timur, Indonesia
*Corresponding author, e-mail: lailiyah@gmail.com[1], ekawicida@gmail.com[2], reza.andrea@gmail.com[3]

Abstract

The inability of students to absorb the knowledge conveyed by the teacher is'nt caused by the inability of understanding and by the teacher which isn't able to teach too, but because of the mismatch of learning styles between students and teachers, so that students feel uncomfortable in learning to a particular teacher. It also happens in senior high school (SHS/SMAN) 1 Anggana, so it is necessary to do this research, to analyze cluster (group) of student learning style by applying data mining method that is k-Means and Fuzzy C-Means. The purpose was to know the effectiveness of this learning style cluster on the development of absorptive power and improving student achievement. The method used to cluster the learning style with data mining process starts from the data cleaning stage, data selection, data transformation, data mining, pattern evolution, and knowledge development.

Keywords*: fuzzy C-Means, K-Mean clustering, learning style*

1. Introduction

The concept of learning according to The United Nations Educational, Scientific and Cultural Organization (UNESCO), requires every educational unit to be able to develop four pillars of education both for now and the future, namely: learning to know, learning to do in this case learners are required to be skilled in doing something, learning to be, and learning to live together. Learning is the means by which a person acquires and develops new knowledge, skills, capabilities, behaviours and attitudes, who is unskilled to be skilled, who does not know how to do something to be able to do something that is all the result of experience or interactively with the environment that performed deliberately. Thus, change that occurs in learners is the process of learning in other words called learning outcomes. Experts in the field of education find the fact that each individual student has a learning style.

A further development of group analysis is to consider the level of membership that includes the fuzzy set as a weighting basis for a grouping called fuzzy clustering [1]. This method represents the development of a strict partition method (K-Means) by doing fuzzy weighting that allows objects to be able to join each group. One technique that is part of the nonhierarchical method is to use Fuzzy C-Means logic (FCM). This algorithm was first introduced by Dunn in 1974. In general, the FCM algorithm is based on the objective function derived from the distance calculation of the center of the group [2]. With this technique the object will tend to belong to a group where the object has the highest degree of membership to the group.

Data mining and cluster analysis research that discusses these 2 methods has been done [1-6] for education [2, 7-9], health [3, 4, 10, 11], and others [12, 13]. Some studies have combined it with the big data theory [14-16] or hybrid theory [17, 18]. Cluster analysis is applied in education with the purpose to improving the teaching and learning process. Contribution of this research is innovation of teaching methods with data mining theory. In this research would be performed data processing of student learning style in senior high school (SHS/SMAN) 1 Anggana with method of data mining K-Means and FCM with the aim to give good and effective learning process. The classification was performed into student clusters and determines appropriate learning method decisions on these groups. The final result of this study was expected to improve students' ability to absorb knowledge from teachers.The objectives of this study are to cluster the learning style with two methods of data mining, compare and analyz the

Received February 26, 2018; Revised January 30, 2019; Accepted February 28, 2019

groups of learning style outcomes of K-Means and Fuzzy C-Means methods, and formulate appropriate learning style decisions for each class of students.

2. Related Works

Research on the game with the same technique has been widely done among others:

- Comparative Analysis of K-Means and Fuzzy C-Means Algorithms [1].
- K-Means Cluster Analysis for Students Graduation: Case Study: STMIK Widya Cipta Dharma [6].
- Application of learning analytics using clustering data Mining for Students' disposition analysis [7].
- Impact of Distance Metrics on The Performance of K-Means and Fuzzy C-Means Clustering an Approach to Assess Student's Performance In E-Learning Environment [8].
- Cluster Analysis for Learning Style of Vocational High School Student Using K-Means and Fuzzy C-Means (FCM) [9].
- Comparative Study of K-Means and Fuzzy C-Means Algorithms on the Breast Cancer Data [10].
- Performance Assessment of K-Means, FCM, ARKFCM and PSO Segmentation Algorithms for MR Brain Tumour Images [11].

In a study conducted by Ghosh and Dubey [1], the comparative K-Means and FCM algorithms were measured by looking at the iteration of centroid point movement. This study looked at the accuracy and weaknesses of both methods in solving the clustering problems in some experimental cases. Research conducted by Wijayanti, et.al [6], their study examines the comparative application of methods K-Means in a case study, namely graduate student grouping for academy (STMIK) Widya Cipta Dharma based on the characteristics of the GPA, the study period, Department of Study Programs and Predicate. Determination of the number of groups is done through a validity index.

In the Bharara's team research [7], the main objective of their research work is to find meaningful indicators or metrics in a learning context and to study the inter-relationships between these metrics using the concepts of learning analytics and educational data mining, thereby, analyzing the effects of different features on student's performance using disposition analysis. Their project, K-Means clustering data mining technique is used to obtain clusters which are further mapped to find the important features of a learning context. Relationships between these features are identified to assess the student's performance.

Research comparing these two methods was also carried out by several researchers. Mahatme, et.al [8], their study helps the researchers to take quick decision about choice of metric for clustering. In clustering algorithm, distance metrics is a key constitute in finding regularities in the data objects. In this paper, impact of three different metrics Euclidean, Manhattan and Pearson correlation coefficient on the performance of K-Means and fuzzy C-Means clustering is presented. In clustering, detection of similarity using distance metrics affects the accuracy of the algorithm [8]. The other case studies, Dubey, et.al [10] also compare these methods. The two main objectives of their work were: firstly, to compare the performance of K-Means and fuzzy C-Means (FCM) clustering algorithms; and secondly, to make an attempt to carefully consider and examine, from multiple points of view, the combination of different computational measures for K-Means and FCM algorithms for a potential to achieve better clustering accuracy. The computational results indicate that FCM algorithm was found to be prominent and consistent than K-Means algorithm when executed with different iterations, fuzziness values, and termination criteria. It is more potentially capable in classifying breast cancer Wisconsin dataset as the classification accuracy is more important than time.

Still in the health topic, in Karegowda's team research [11], they compare the performance of K-Means, Fuzzy C-Means (FCM), Particle Swarm Optimisation (PSO) and Adaptive Regularised Kernel Fuzzy C-Means (ARKFCM)-based segmentation techniques for accurate delineation of tumour using clinical brain tumour Magnetic Resonance images. Their experimental evaluation revealed K-Means and FCM segmentation algorithms out performed compared with PSO and ARKFCM segmentation algorithms. Andrea, et.al [9], their research is similar to this research, they analyze cluster (group) type of student learning by applying K-Means and Fuzzy C-Means (FCM), but their paper case study is High School Student Penajam Paser Utara. The differences in this research are emphasized on the application of

K-Means and FCM methods for clustering student learning style on Senior High School (SHS/SMA), as well as measuring the level of validity of the final model of each method. The results of this study can be used by the school to help in taking policy to determine the appropriate teaching model in each class.

3. Research Stages

The research method used was experiment with research stages as follows:

- Data collection
  Collecting questionnaire data from 100 students
- Preliminary data processing (data cleaning)
  The collected data was processed by soft-computing algorithm to reduce irrelevant data. While relevant data and analysis tasks were returned into the database (selection process).
- Formation of proposed model (data transformation)
  In this method, data mining would be described schematically and accompanied by a calculation formula. The model would be formed from the data that already processed. The result of model processing would be measured with the current model.
- Experiments and Model Testing
  Describes how experiments were carried out until the formation of the model and explains how to test the model that was formed.
- Evaluation and validation of results (pattern evaluation)
  The evaluation was performed by observing the cluster results with both soft-computing algorithms. Validation was performed by measuring the cluster results and compared with the original data. Performance measurement was performed by comparing the error value of cluster result of each algorithm so that it can be known more accurate algorithm.
- Knowledge presentation
  An overview of visualization and knowledge techniques was used to provide knowledge to users. At this stage the development of knowledge was used by the school to take policy in determining the appropriate teaching model in school.

4. Data Collection

The collected data consists of secondary data and primary data. Primary data directly from questionnaires and interviews in SHS 1 Anggana. While the secondary data was obtained by studying literature studies in the form of written rules or documents that have relation to the title research. In addition, the data was obtained through observation or direct observation of conditions in the field that is in the environment of SHS 1 Anggana.

5. Research Methods

After the data was collected then the next stage was to prepare the data in order to be used for data mining process. The raw data can be used for the data mining process. The raw data to be used in this application was obtained from the questionnaire of 100 students. Preliminary data processing is part of the data preparation. The steps taken include eliminating the double data and cleaning the data that was plagued, combine the data, determined the attribute to be processed and change the data. Data preparation was performed manually by using excel format *.csv. The result of data preparation process was presented in tabular form Table 1.

Student data in SMAN 1 Anggana based on the type of learning questionnaire that was filled with 100 students of random samples from classes of 1, 2 and 3 from various departments. Where: X_1 is the percentage of learning style with visual learning; X_2 is the percentage of auditory learning; X_3 is the percentage of kinesthetic learning. Data from Table 1 can be grouped into several groups according to the attributes that have been determined in the form of X_1 (Visual), X_2 (Auditory), X_3 (Kinesthetic).

Table 1. Data of Student Learning Styles Questionnaire

	X_1 (Visual)	X_2 (Auditory)	X_3 (Kinesthetic)
1	26.66667	40	33.33333
2	46.66667	53.33333	0
3	26.66667	60	13.33333
4	26.66667	46.66667	26.66667
5	46.66667	40	13.33333
6	33.33333	53.33333	13.33333
7	20	60	20
8	33.33333	53.33333	13.33333
..	..	..	..
100	26.66667	53.33333	20

5.1. K-Means Algorithm

K-Means was first published by Stuart Lloyd in 1984 and is a widely used clustering algorithm. K-Means works by segmenting existing objects into clusters or so-called segments so that objects within each group are more similar to each other than objects in different groups. The clustering algorithm is putting a similar value in one segment, and putting different values in different clusters [19]. K-Means separates data optimally with a loop that maximizes the result of the partition until no data changes in each segment. K-Means works with a top-down approach because it starts with pre-defined segmentation [20]. So the result of data of a segment is not possible mixed between one segments with other segment [21]. This approach also speeds up the computation process for large amounts of data.

The K-Means algorithm applies to objects represented in d-dimensional vector dots. K-Means clustered all the data in each dimension where the point in the same segmentation was given cluster ID. The value of *k* is the basic input of the algorithm that determines the number of segments to be formed. Partition will be formed from a set of object n into cluster *k* so as to form the similarity of object in each *k*-segmentation. The K-Means algorithm is a widely used algorithm for determining clusters [22], because it is easy to use, has exact and modifiable calculations to meet the needs of use.

5.2. Fuzzy C-Means Algorithm

The famous fuzzy clustering algorithm is FCM introduced by Jim Bezdek. He introduced the idea of the fuzzification parameters (m) within the range [1, n] that determines the fuzzy degree of the cluster. When cluster m=1, the effect is a clustering crips from some point, but when m>1 the fuzzy degree between points in the decision space becomes increased [20, 23]. FCM clustering involves two processes: the calculation of the cluster center and the mastery of the point toward the center by using a form of Euclidean distance. This process is repeated until the center of the cluster has stabilized. FCM executes a direct constraint of the fuzzy membership function connected to each point. The purpose of the FCM algorithm is the assignment of data points into clusters with varying degrees of membership. This membership reflects the degree to which points are more representative of one cluster [24].

6. Results Analysis

Find the right number of groups with optimal cluster number recommendations can be seen on the evacluster chart. The evalcluster graph is the best recommendation graph for group assignment, which will be used for grouping of data. The first highest peak of evalclusters chart will be used for the best cluster determination of some existing clusters [25], the best cluster according to the evalcusters for the above data is in cluster 4 of 96.863 that shown on Figure 1. Based on the cluster formed in Table 2, type of student learning in SHS 1 Anggana can be grouped into four groups according to the values that meet on each variable in each cluster and can be seen in the silhouette of cluster 4 in Figure 2. Figure 2 shows, 4 clusters of the silhouette image. It can be seen that very few cluster elements were in negative territory. Thus the result of this cluster was quite good and represents similar groups.

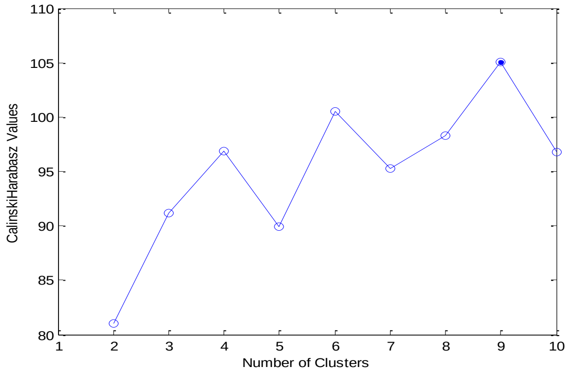


Figure 1. Evacluster graph

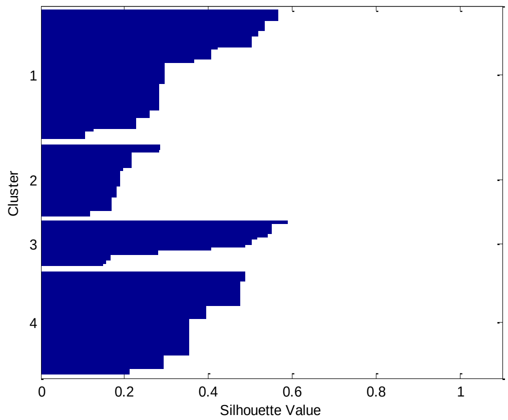


Figure 2. Silhouette with 4 clusters

Table 2. Iteration Data of Each Cluster

		2 Clusters		
	iter	phase	num	Sum
iteration	1	1	132	33039.1
	2	1	10	32004.6
	3	1	13	30783.9
	4	2	1	30616.1
	5	2	0	30555.2
		Best total sum of distances=30555.2		
Centroid	50.8642	32.4691	16.6667	
	32.1368	47.7778	20.0855	
		3 Clusters		
	iter	phase	num	Sum
iteration	1	1	132	27031.9
	2	1	13	24736
	3	1	13	22503.3
	4	1	1	22341.2
	5	2	0	22341.2
		Best total sum of distances=22341.2		
Centroid	51.1565	31.0204	17.8231	
	26.8571	44.3810	28.7619	
	37.6389	50.1389	12.2222	
		4 Clusters		
	iter	phase	num	Sum
iteration	1	1	132	19418.9
	2	1	9	17631.1
	3	1	4	17253.7
	4	2	0	17253.7
		Best total sum of distances=17253.7		
Centroid	26.6667	46.6667	20	
	53.3333	46.6667	6.66667	
	40	26.6667	33.3333	
	46.6667	40	13.3333	

6.1. K-Means Cluster Analysis

The process of centroid deployment into 4 clusters by using a 3D graph that compares the attributes used, shown in Figure 3. Figure 3 shows, obtained the percentage value of 100 student samples: Cluster 1: 37%; Cluster 2: 20%; Cluster 3: 13%; Cluster 4: 30%. Then K-Means analysis can be drawn:

a. 37% of students have auditory learning dominant only.
b. 20% of students have visual learning style and little audio help (mixed visual-auditory)
c. 13% of students have balanced blend of the three styles
d. 30% of students are like cluster 3, they have visual learning style and little audio help (mixed visual-auditory), but this cluster has kinesthetic point more than kinesthetic point in cluster 3, and visual-auditory point less than visual-auditory point in cluster 3.

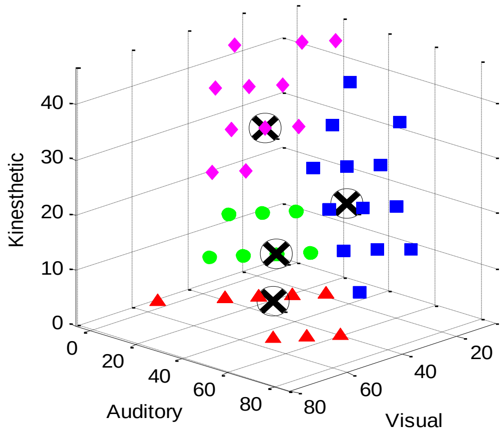


Figure 3. Surface 3D graph of 4 K-Means clusters

6.2. FCM Analysis

The FCM grouping was conducted with the same group as the optimal group of K-Means clusters (4 clusters), in order to compare the results of cluster patterns formed. From Figure 4, it was obtained FCM algorithm on 4 clusters showed that the clustering process stopped at 100th iteration with the objective function value was 0.8977×10^4. The number of iterations of 4 clusters was less and effective compared to 5 or 6 clusters. Figure 5 shows, obtained the percentage value of 100 student samples: Cluster 1: 21%; Cluster 2: 33%; Cluster 3: 24%; Cluster 4: 22%. Then FCM analysis can be drawn:
- 21% of students have auditory learning dominant only.
- 33% of students have auditory learning style and little visualization help (mixed auditory-visual)
- 24% of students have balanced blend of the three styles
- 22% of students have visual learning style and little audio help (mixed visual-auditory)

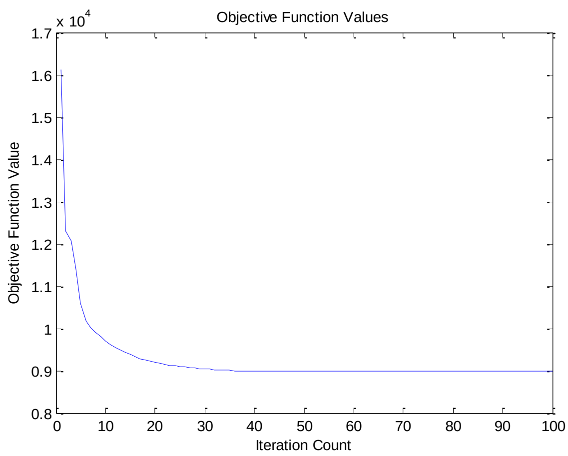


Figure 4. Graph of objective function values on 4 clusters

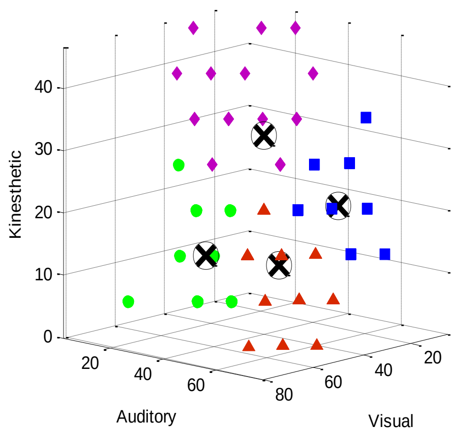
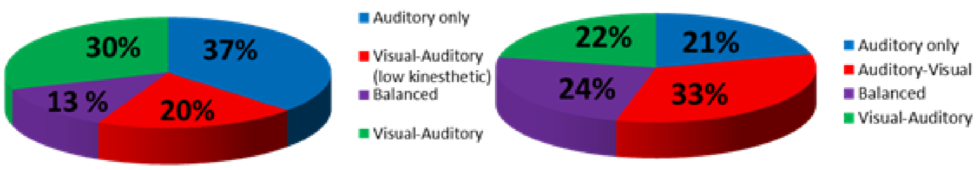


Figure 5. Surface 3D graph of 4 FCM Clusters

6.3. Comparison of K-Mean and FCM Analysis

Both algorithms resulted in nearly identical clustering of 4 clusters, and with numbers that had a small percentage increment. The two percentages of clusters can be seen in Figure 6. Figure 6, the highest percentage of auditory learning, and the second highest is mixed auditory-visual or visual-auditory, while the low percentage is in balanced blend of the three styles. But there is little difference in the results of K-Means and FCM cluster analysis that is in the 2nd cluster. K-Means analyzed that the 2nd cluster was a group of students have visual learning more dominant than auditory (mixed visual-auditory), while FCM analyzed that the 2nd cluster was a group of students have auditory learning more dominant than auditory (mixed visual-auditory).



(a) (b)

Figure 6. Graph of learning style grouping percentage with (a) K-Means and (b) FCM

7. Conclusion

From analysis results of the two cluster algorithms used can be drawn conclusion: The classification of learning style of high school students SHS 1 Anggana by using K-Means and FCM can be formed into 4 clusters. Many students of SHS 1 Anggana liked to learn with auditory learning, that assisted with visualization rather than learning just by reading or self-practice. This conclusion is drawn from the merging of clusters percentage of students who favor mixed auditory–visual learning plus the percentages of who only favor auditory learning. This research can help the teachers of SHS 1 Anggana to find the right method of teaching to their students in class.

References

[1] Ghosh S, Dubey SK. Comparative Analysis of K-Means and Fuzzy C-Means Algorithms. *International Journal of Advanced Computer Science and Applications*. 2013; 4(4): 35-39.

[2] Wijayanti S, Andrea R. *K-Means Cluster Analysis for Students Graduation: Case Study: STMIK Widya Cipta Dharma*. In Proceedings of the 2017 International Conference on E-commerce, E-Business and E-Government. ACM. 2017: 20-23.

[3] Cebeci Z, Yildiz. Comparison of K-Means and Fuzzy C-Means Algorithms on Different Cluster Structures. *Agrárinformatika/journal of agricultural informatics*, 2015; 6(3): 13-23.

[4] Mane DS, Gite BB. Brain Tumor Segmentation Using Fuzzy C-Means and K-Means Clustering and Its Area Calculation and Disease Prediction Using Naive-Bayes Algorithm. *Brain*, 2017; 6(11): 21342-21347.

[5] Modi H, Baraiya N, Patel H. Comparative Analysis of Segmentation of Tumor from Brain MRI Images Using Fuzzy C-Means and K-Means . *Fuzzy Systems*. 2018; 10(1): 14-18.

[6] Nasir ASA, Jaafar H, Mustafa WAW, Mohamed Z. *The Cascaded Enhanced K-Means and Fuzzy C-Means Clustering Algorithms for Automated Segmentation of Malaria Parasites*. In MATEC Web of Conferences EDP Sciences. 2018; 150: 06037.

[7] Bharara S, Sabitha S, Bansal A. Application of learning analytics using clustering data mining for Students' disposition analysis. *Education and Information Technologies*, 2018; 23(2): 957-984.

[8] Mahatme VP, Bhoyar KK. Impact Of Distance Metrics on The Performance of K-Means And Fuzzy C-Means Clustering-An Approach To Assess Student's Performance In E-Learning Environment. *International Journal of Advanced Research in Computer Science*, 2018; 9(1): 887-892.

[9] Andrea R, Palupi S, Qomariah S. Cluster Analysis for Learning Style of Vocational High School Student Using K-Means and Fuzzy C-Means (FCM). *Jurnal Penelitian Pos dan Informatika*, 2017; 7(2): 121-128.

[10] Dubey AK, Gupta U, Jain S. Comparative Study of K-Means and Fuzzy C-Means Algorithms on The Breast Cancer Data. *International Journal on Advanced Science, Engineering and Information Technology*. 2018; 8(1): 18-29.

[11] Karegowda AG, Poornima D, Sindhu N, Bharathi PT. Performance Assessment of K-Means , FCM, ARKFCM and PSO Segmentation Algorithms for MR Brain Tumour Images. *International Journal of Data Mining and Emerging Technologies*. 2018; 8(1): 18-26.

[12] Buditjahjanto IA, Miyauchi H. An intelligent decision support based on a subtractive clustering and fuzzy inference system for multiobjective optimization problem in serious game. *International Journal of Information Technology & Decision Making.* 2011; 10(05): 793-810.
[13] Krinidis S, Chatzis V. A robust fuzzy local information C-Means clustering algorithm. *IEEE transactions on image processing*, 2010; 19(5): 1328-1337.
[14] Hassani H, Silva ES. Forecasting with big data: A review. *Annals of Data Science*, 2015; 2(1): 5-19.
[15] Cai X, Nie F, Huang H. Multi-View K-Means Clustering on Big Data. In *IJCAI*. 2013: 2598-2604.
[16] Shirkhorshidi AS, Aghabozorgi S, Wah, TY, Herawan, T. *Big data clustering: a review*. In International Conference on Computational Science and Its Applications. Springer, Cham. 2014: 707-720.
[17] Cheng D, Ding X, Zeng J, Yang N. Hybrid K-Means Algorithm and Genetic Algorithm for Cluster Analysis. *Indonesian Journal of Electrical Engineering and Computer Science*. 2014; 12(4): 2924-2935.
[18] Mahboub A, Arioua M. Energy-efficient hybrid K-Means algorithm for clustered wireless sensor networks. *International Journal of Electrical and Computer Engineering (IJECE)*. 2017; 7(4): 2054-2060.
[19] Li D, Wang S, Li D. Spatial data mining. Springer Berlin Heidelberg. 2015.
[20] Witten IH, Frank E, Hall MA. DataMining Practical Machine Learning Tool and Techniques. 3rd. *Elsevier Inc*. USA. 2011.
[21] Aggarwal CC, Reddy CK. (Eds.). Data clustering: algorithms and applications. CRC press. 2013.
[22] Han J, Pei J, Kamber M. Data mining: concepts and techniques. Elsevier. 2011.
[23] Dean J. Big Data, Data Mining, and Machine Learning Value Creation for Business Leaders and Practitioners. Wiley. New Jersey. 2014.
[24] Taher A. Adaptive neuro-fuzzy systems. In Fuzzy systems. Intech. 2010.
[25] Ledolter J. Data Mining and Business Analytics with R. Wiley. New Jersey. 2013.