

Data Distribution Aware Classification Algorithm based on K-Means

Tamer Tulgar

Department of Computer Engineering
Girne American University
Girne, T.R.N.C.
Mersin 10 Turkey

Ali Haydar

Department of Computer Engineering
Girne American University
Girne, T.R.N.C.
Mersin 10 Turkey

İbrahim Erşan

Department of Computer Engineering
Girne American University
Girne, T.R.N.C.
Mersin 10 Turkey

Abstract—Giving data driven decisions based on precise data analysis is widely required by different businesses. For this purpose many different data mining strategies exist. Nevertheless, existing strategies need attention by researchers so that they can be adapted to the modern data analysis needs. One of the popular algorithms is K-Means. This paper proposes a novel improvement to the classical K-Means classification algorithm. It is known that data characteristics like data distribution, high-dimensionality, the size, the sparseness of the data, etc. have a great impact on the success of the K-Means clustering, which directly affects the accuracy of classification. In this study, the K-Means algorithm was modified to remedy the algorithm's classification accuracy degradation, which is observed when the data distribution is not suitable to be clustered by data centroids, where each centroid is represented by a single mean. Specifically, this paper proposes to intelligently include the effect of variance based on the detected data distribution nature of the data. To see the performance improvement of the proposed method, several experiments were carried out using different real datasets. The presented results, which are achieved after extensive experiments, prove that the proposed algorithm improves the classification accuracy of K-Means. The achieved performance was also compared against several recent classification studies which are based on different classification schemes.

Keywords—Classification; k-means; variance effect; big data

I. INTRODUCTION

Data Mining can be defined as the area of information science which analyses raw data to produce meaningful information by extracting useful patterns [1]. Because of this nature, Data Mining has been among the vital information processing tools [2].

During the past decade, the nature of data changed drastically. Today, businesses tend to make decisions based on data driven analysis [3]. To achieve more precise decision making, businesses need to analyse data coming from several resources including popular social media sources, digital data warehouses, cloud storages, etc. using many different sources results in highly unstructured and vast volumes of data. Today, during the Big Data Era, the classical data analysis techniques need to change and improve to cope with the continuously increasing velocity, variety and the volume of the data which needs to be analysed [4].

One of the most important data mining tasks is the classification. Classification, which is the task of assigning objects to one of several predefined categories, is a pervasive problem

that encompasses many diverse applications [1]. Recently, many research studies [5]–[15] are carried out to improve the performance and solve the shortcomings of several known data classification algorithms so that the modern data analysis needs can be met.

One of the important challenges that needs to be addressed in classification is correctly grouping the related data in correct clusters, especially when the data is radically distributed. Classically, there are many well accepted classification methods [16]. K-means [17] is one of the most famous partition clustering algorithms because it is a very simple, statistical and a quite scalable method [18].

Nevertheless, just like other classical classification algorithms, to apply K-Means in today's data mining tasks, the algorithm needs to be adapted to cope with unstructured, highly dimensional data and when the distribution is not suitable to be successfully clustered by data centroids, where each centroid is represented by a single mean [18].

K-Means is a prototype-based, partitional clustering technique that attempts to find a user-specified number of clusters (K), which are represented by their centroids [1]. To summarise the clustering process of K-Means: First K random instances from the data set are chosen and the other instances of the data set are grouped around the randomly chosen K centroids according to their proximity or similarity of the centroids. Then, the means of the formed clusters are calculated and become the new centroids. Afterwards, re-grouping is performed according to the newly found centroids. This process continues iteratively until the calculated means of the clusters do not vary anymore [17]. This process is called training the algorithm. To perform classification, a new data instance is compared against the formed centroids of the data during the training phase and the classification decision is based on the minimum proximity of the data instance to the cluster centroids.

Hence, for the K-Means algorithm, the success of the classification decision can be expressed as how accurately the new instance was classified to the correct cluster and strongly depends on the training success. The success of the training can be detected by using a well selected validation data or by cross-validation [18].

In this paper, an improved data distribution aware K-Means algorithm is proposed to improve the classification accuracy when K-Means fails to successfully classify data under varying data distributions in datasets. The proposed improvement is

mainly introducing the effect of variance to the classification decision so that the tested data instance can be more precisely classified under conditions which are otherwise challenging for the classical K-Means algorithm.

To evaluate the performance of the proposed algorithm, extensive experiments are carried out using several real datasets. The results achieved after the experiments prove that, the proposed algorithm improves the K-Means algorithm.

The rest of this paper is organised as follows: Section 2 summarizes some related studies. In Section 3, the proposed method is explained focusing on how K-Means algorithm is improved. The experimentation method, the real datasets used during experiments are given in Section 4. The achieved performance results and comparisons with other algorithms are discussed in Section 5. Finally, Section 6 concludes the paper.

II. SOME RECENT LITERATURE ON CLASSIFICATION

In [9], three schemes for classification are proposed and compared. The proposed schemes are K-nearest Neighbour (KNN), Fuzzy KNN and the Support Vector Machine (SVM). The proposed schemes are applied as a part of MapReduce system [19]. The Fuzzy KNN proposed in [9] employs Gaussian Membership Functions as the representatives of the data clusters, which is one of the details pointed out in [9]. In the results the author presents the experimental results which show that among their proposed alternatives the scheme which combines Support Vector Machine with Soft Labels produces the better classification accuracies.

Another MapReduce fuzzy data classification scheme is proposed in [10]. In [10], the authors propose four different schemes and compare their performances. The four proposed classification techniques are fuzzy KNN and mode function, SVM classifier and mode function, SVM and soft labels and finally SVM classifier and fuzzy Gaussian membership function. In [10], the four methods mainly differ in the Reducer function part of the MapReduce such that the reducers are implemented using three approaches which are, the mode, the soft labels and fuzzy Gaussian. The results presented in the study illustrate that the fuzzy techniques perform better than the crisp methods. Especially, the SVM using soft labels produces the better results.

The study presented in [11] investigates the efficiency of Gaussian Mixture Models (GMM) and fuzzy Expectation Maximisation (EM). The technique proposed in [11] mainly focuses on clustering and classification of fuzzy data. The results presented in [11] illustrate that the proposed technique is a contribution which helps estimating the distribution of imprecisely known data. The authors also claim to improve the classification accuracy of noisy data, which they present in their results.

In [12], a K-Means variation together with a KNN classification approach is proposed by the authors. The proposed method in [12], clusters the data using the K-Means algorithm and then for testing relies on KNN Classification. It is claimed by the authors of [12] that their proposed method is suitable for dealing with big data. The results they present outperforms the results of [13], which will be summarized next in this section.

The method proposed in [13] modifies the KNN algorithm with a self representation of the data clusters ideology. The presented main aim is to learn an optimal k value in KNN to improve the accuracy of the classification. To support their claim, the authors compare their results with three other algorithms named as kNNC, LMMN and ADNN which are summarized in [13]. The results presented in the paper shows better performance when compared to these three algorithms.

Authors of [14] compare and analyse five different existing methods to deduce the strengths and weaknesses of the KNN classification scheme for big data. As evaluation, [14] presents the advantages and disadvantages of the different stages of the compared classification models which are all applied on MapReduce work-flow. It is claimed in [14] that the results achieved in the study can be used to tackle different practical KNN problems in the context of big data.

In [15], another KNN based classification scheme is proposed. The proposed study in [15] can be mainly summarised as an iterative version of MapReduce work-flow based on SPARK which benefits from the KNN classification. The performance of the method proposed in [15] is evaluated using experiments. The results of the presented experiments illustrate that the method performs better than the KNN approaches based on Hadoop both from accuracy and runtime points of view.

III. PROPOSED VARIANCE IMPROVED K-MEANS ALGORITHM

The proposed method presented in this paper can be summarized as an improved K-Means algorithm which can tackle with close centroids of different classes with different variances that can be seen in different datasets.

The main ideology of the contribution is an *a priori* decision that will detect whether the effect of variance of the data to be classified should be taken into consideration or not. It is shown in Section 5 that the improvement expectation of the proposed algorithm is met and is visible in the comparative results.

In this section, before explaining the main contribution of the proposed work in detail, first the classical K-Means algorithm will be summarised so that the nature of the whole classification scheme can be better understood.

A. Overview of the Classical K-Means Algorithm

The K-Means is a classical prototype-based, partitional clustering technique which tries to cluster the given data into user specified K-clusters [17].

Typically, any dataset to be clustered will be containing elements, which will be called instances in the algorithms hereafter. The instances of a dataset will have class labels which identify their belonging information. For example, the physical features of human beings will result in classifying the humans into man, women and children classes.

The K-Means algorithm will use the features of instances in a dataset and try to cluster the instances of the classes of the dataset into K number of clusters. Clustering performed by K-Means can be summarized as follows:

The algorithm first chooses K number of random instances from each class of the dataset. Using these randomly chosen instances as the initial centroids, K-Means measures the euclidean distances of the instances to the centroids.

By considering the minimum distance as the objective, K-Means forms K number of groups of instances. Afterwards, the mean values (μ) of the K groups will be calculated. The calculated K number of means become the new centroids of the K instance groups. The mean(μ) of the instances are calculated by the well known mean value formula shown in the following equation, where X_i represents the instance in a group and the N is the size of the formed group.

$$\mu = \frac{1}{N} \sum_{i=1}^N X_i$$

Next, the instances of the classes will be re-grouped according to the minimum distances of the instances to the new centroids. The means of the K new groups will be calculated and the K means of the K groups will become the new centroids. In the equation shown below, $\|X_i - \mu_j\|^2$ demonstrates the distance of the instance X_i to the centroid μ_j .

Hence, the membership of an instance X_i is decided based on the minimum T of all centroids μ_j .

$$T = \min_j (\|X_i - \mu_j\|^2), \forall j = 1..k$$

This iterative process continues until the previous and the new centroids are the same.

When the process stops, the centroids represent the K clusters formed by the K-means clustering algorithm. The whole algorithm is illustrated in Fig. 1.

The whole clustering process explained above is actually the training phase of a classification task. In classification using K-Means, when a new data instance is needed to be identified (i.e. when the data class of the instance needs to be detected from the features of the instance), the euclidean distance of the instance is measured to the final centroids of the previously formed clusters and the class of the centroid producing the minimum distance is identified as the class of the newly arriving data instance.

It is known that the classical K-Means algorithm tends to form proper clusters when the resulting clusters of a dataset are relatively uniform in size [18]. In the contrary case, when the formed clusters are used in classification, ambiguities in membership decisions can exist. In other words, when the distances of the new data is similarly close to several centroids, wrong identifications may become possible.

The algorithm presented in this paper, tries to remedy the ambiguity explained above so that the K-Means classification accuracy can be improved.

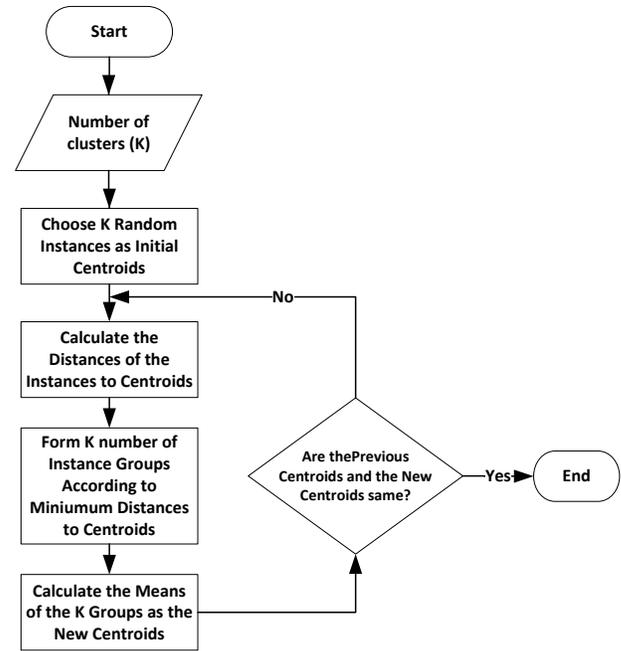


Fig. 1. The K-Means algorithm.

B. Proposed K-Means-Mod Algorithm

To improve the classification accuracy of the K-Means algorithm under different data distributions, the effect of variance is included in the proposed modified K-Means algorithm, which will be referred to as K-Means-Mod during the rest of this paper.

K-Means-Mod first decides if the effect of variance of the data should be considered or not at the end of the training phase according to the nature of the formed clusters. Later, the algorithm bases its testing phase on the previous variance usage decision.

The variance can be defined as the measure of how spread out the distribution of a group of data is [20]. Variance can be defined with the following equation, where X_i is the instance and μ_j is the centroid of the cluster to be tested.

$$\sigma_j^2 = \frac{1}{N} \sum_{i=1}^N (X_i - \mu_j)^2$$

In classification using K-Means, when a new data instance will be tested against the formed clusters, the distance of the instance to more than one clusters can be similarly close. The possibility of this case is higher when different datasets contain instances which are not clearly different from each other. These kind of clusters are frequently seen when the dataset to be clustered contains populations which may not be successfully clustered by centroids calculated using single mean representation.

Since the K-Means algorithm assumes that the formed clusters are clear partitions of the whole data that are tightly grouped together, classical K-Means only relies on the minimum distance as the decision criteria.

To compensate the distance calculation based decisions of K-Means, the proposed K-Means-Mod algorithm includes the effect of how far the instances are spread around the centroids to the classification decision by including the variance to the decision.

This compensation is achieved by dividing the calculated distances to the clusters to the variance of the tested element and the cluster members. This results in concluding that a cluster with a smaller variance will produce a stronger membership strength versus a cluster with a greater variance will produce a weaker membership strength. Hence, the membership strength measure is defined with the following equation:

$$\text{membership strength} = \frac{\text{dist}(X_i, \mu_j)}{\sigma_j^2}$$

The proposed K-Means-Mod algorithm, bases its classification decision on either the membership strength or classical distance measurement according to the variance calculations after training.

The decision is given by validating the formed clusters' correctness by the accuracy performance of classifying the training data.

The part of the algorithm which starts after the training phase is illustrated in the flowchart which is presented in Fig. 2.

With this compensation idea, the proposed K-Means-Mod algorithm either performs similar to classical K-Means or when K-Means is misled by the data distribution, better than the classical K-Means accuracy.

IV. PERFORMANCE ANALYSIS

To evaluate the classification accuracy of the proposed K-Means-Mod algorithm, classification experiments are conducted using real datasets downloaded from UCI Machine Learning Repository [21].

A. Used Datasets

For the experiments six real datasets are used in the experiments which are summarised in the following table (Table I):

TABLE I. DATASETS USED IN THE EXPERIMENTS

Dataset	Instances	Features	Classes
ionosphere	351	34	2
wdbc	569	32	2
seeds	210	7	3
wine	178	13	3
satimage	6435	36	7
pendigits	10992	16	10

1) *Ionosphere*: Ionosphere data set is the data coming from the classification of radar returns from the ionosphere. The dataset contains 351 instances belonging to 2 classes. Each instance contains values belonging to 34 features. This dataset is also used in [13].

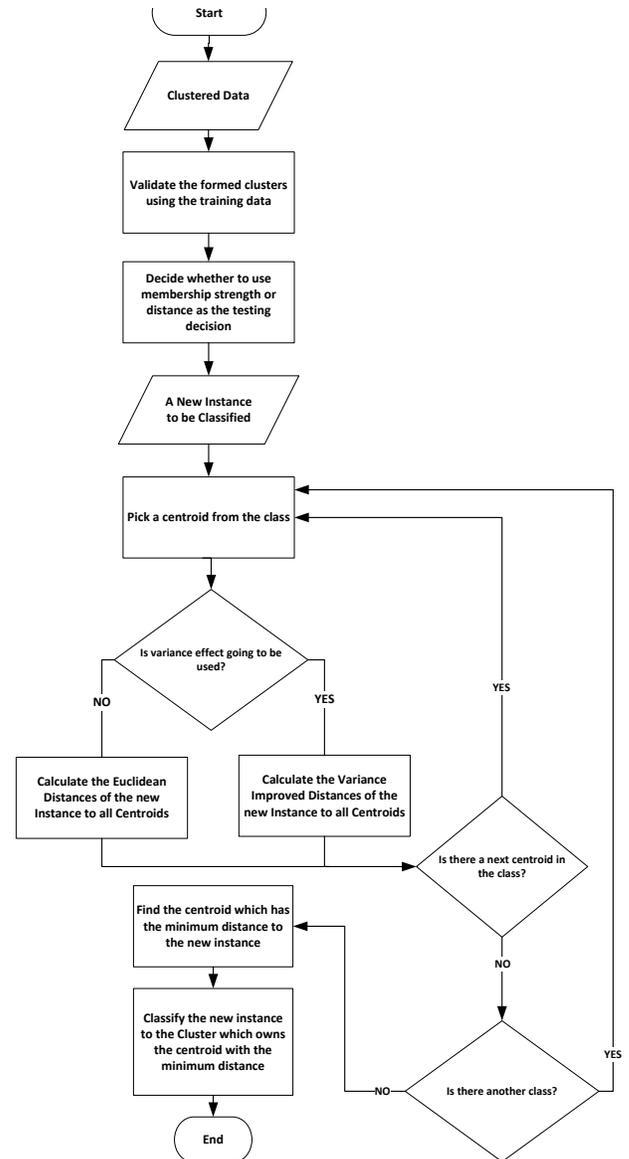


Fig. 2. K-Means-Mod algorithm.

2) *WDBC*: The Wisconsin Diagnostic Breast Cancer (WDBC) was first used in [22]. The dataset contains 569 instances belonging to 2 classes. Each instance contains values belonging to 32 features. WDBC dataset is also used in [13].

3) *Seeds*: The seeds dataset contains the measurements of geometrical properties of kernels belonging to three different varieties of wheat. The dataset contains 210 instances in 3 classes. Each instance is defined by the values of 7 features. Seeds data set is first used in [23] and also investigated in [13].

4) *Wine*: Wine dataset contains data from chemical analysis to determine the origin of wines. The dataset is composed of 178 instances in 3 classes containing 13 features. Wine dataset is also used in the experiments of [13].

5) *Satimage*: The Satimage dataset was generated from Landsat Multi-Spectral Scanner image data. The dataset contains 6435 instances belonging to 7 classes. Each instance contains the data of 36 features. Satimage dataset is also used

by [11]–[13].

6) *Pendigits*: Pen-Based Recognition of Handwritten Digits Data Set (pendigits) is a digit database of 250 samples from 44 writers [24]. This dataset contains 10992 instances belonging to 10 classes. Each instance contains the data of 16 features. Pendigits is also used by [11]–[13].

B. Experimental Setting

The K-Means and the K-Means-Mod algorithms are coded in JAVA language [25]. Experiments were executed on a Core i7 CPU with 16 GB Ram PC.

For each dataset used in the experiments, 10-fold cross validation is used and each test is repeated 10 times and the averages of the 10 tests are considered so that the reliable results can be achieved. Reliability of the results achieved after 10 runs is tested by measuring the standard deviation among the achieved results.

To demonstrate the effectiveness of the proposed algorithm, K-Means-Mod was compared against classical K-Means as well as other recent classification schemes.

V. RESULTS

The results presented in this section show the classification accuracy as the performance metric of the proposed algorithm.

The classification accuracy can be defined as follows:

$$\text{Classification accuracy} = \frac{\text{Number of correct class detections}}{\text{Number of total class detections}}$$

As it can be seen in Table II, the proposed K-Means-Mod algorithm improves the accuracy of the classical K-Means in three datasets and performs similarly for the other three datasets.

Also, Table II presents the standard deviations of the 10 runs performed. As it can be seen the repeated experiments resulted in coherent accuracy performance with little deviation among the runs.

TABLE II. K-MEANS VS K-MEANS-MOD ACCURACY(%) PERFORMANCE COMPARISON

Dataset	K-means	K-means-Mod	Std. Dev.
Ionosphere	91.01	91.23	± 1.14
WDBC	94.75	96.28	± 1.02
Seeds	93.24	94.24	± 2.07
Wine	81.33	98.91	± 0.42
Pendigits	97.71	97.70	± 0.06
Satimage	89.80	89.80	± 0.23

The results presented in Fig. 3 show the accuracy performance of the algorithm in six different datasets against the change in the number of centroids (i.e. the K value).

In Fig. 3, it can be observed that when K-Means algorithm demonstrates miss classifications, the proposed K-Means-Mod algorithm significantly performs better in terms of the classification accuracy.

In Table III, the results of the proposed K-Means-Mod algorithm is compared against three other recent classification studies which are [11]–[13]. The three compared studies contain the accuracy performances achieved from fuzzy expectation maximisation and several modified KNN approaches, respectively.

In the comparisons it can be observed that the proposed algorithm performs better than the compared classification algorithms for majority of the tested datasets. For the only dataset WDBC where the proposed K-Means-Mod is not better than its competitor, it is worth noting that the performance of the proposed algorithm and the competitors performances are almost the same with only a 0.22% difference.

Looking at the comparative results it can be seen that, the proposed algorithm decides to classify based on the membership strength or by distance calculations correctly since it does not disturb the K-Means performance when it is on par with the competitor performances.

VI. CONCLUSION

In this paper a decision based, data distribution aware K-Means based classification algorithm is presented and the performance results are compared with several studies.

In the conducted experiments it was observed that the classical K-Means algorithm showed weaknesses in the classification accuracy for some of the datasets analysed. This mainly occurs when the centroids of different classes are very close to each other. This weakness is one of the main drawback when applied to the modern data classification needs.

The proposed contribution to K-Means algorithm detects when this weakness will be experienced and improves the decision making correctness of the K-Means algorithm by introducing a new decision criteria called the membership strength by introducing the effect of the variance to the classification decision.

The presented results show that the proposed contribution practically improves the K-Means algorithms classification accuracy under conditions when K-Means starts failing to correctly classify the data under various data distribution conditions.

Also, the presented comparison results prove that the proposed algorithm performs better than the majority of other approaches recently proposed in the literature and is still a competitor for data classification tasks.

With the achieved results it can be concluded that the well-known K-Means algorithm with the proposed improvement can be usable for the modern data mining needs.

Future Works include adding intelligent feature extraction to the proposed K-Means-Mod algorithm to further improve the classification accuracies as well as the classification delays.

Next, a study will be carried out to apply the proposed K-Means-Mod algorithm to MapReduce workflow to make the algorithm further usable for the modern big data analysis and testing the new scheme for bigger datasets which would be more challenging.

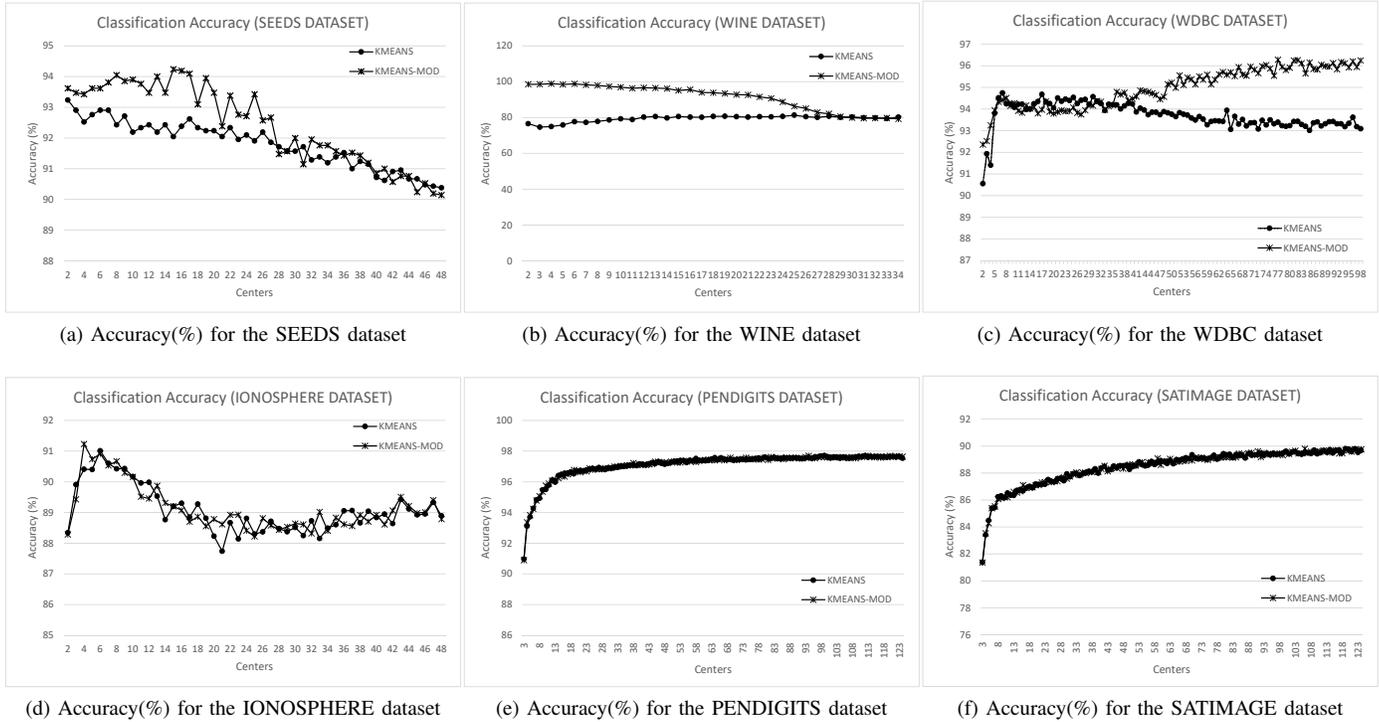


Fig. 3. K-Means vs. K-Means-Mod accuracy comparisons.

TABLE III. K-MEANS-MOD ACCURACY(%) PERFORMANCE COMPARISON

Dataset	Fuzzy-EM [11]	LC-KNN [12]	SR-KNN [13]	K-means-Mod
Ionosphere	-	-	89.71	91.23
WDBC	-	-	96.5	96.28
Seeds	-	-	90.19	94.24
Wine	-	-	97.07	98.91
Pendigits	96.50	97.21	94.52	97.70
Satimage	85.50	88.83	88.06	89.80

Another future work in the project will be to practically test the new scheme in Hadoop and Spark environments on a physical cluster at the Girne American University, Department of Engineering Research Laboratory.

REFERENCES

- [1] Pang-Ning Tan, Michael Steinbach and Vipin Kumar, "Introduction to Data Mining", 1st ed., Reading, MA: Addison-Wesley, 2005.
- [2] Shu-Hsien Liao, Pei-Hui Chu and Pei-Yuan Hsiao, "Data Mining techniques and applications - A decade review from 2000to 2011", Expert Systems with Applications, vol. 39, no. 12, pp. 11303 - 11311, 2012.
- [3] Andrew McAfee and Erik Brynjolfsson, "Big Data: The Management Revolution", Harvard Business Review, October 2012.
- [4] Dilpreet Singh and Chandan K. Reddy, "A survey on platforms for big data analytics", Journal of Big Data vol. 1, no. 8, 2014.
- [5] Adil Fahad et. AL., "A Survey of Clustering Algorithms for Big Data: Taxonomy and Empirical Analysis", IEEE Trans.on Emerging Topics in Computing, vol. 2, no.3, pp. 267-279, 2014.
- [6] Shichao Zhang, Ming Zong and Debo Cheng, "Learning k for KNN Classification", ACM Transactions on Intelligent Systems and Technology, vol. 8, no. 3, pp. 43:1-19, 2017.
- [7] Kun Niu, Fang Zhao and Shubo Zhang, "A Fast Classification Algorithm for Big Data Based on KNN", Journal of Applied Sciences, vol. 13,no. 12, pp. 2208-2212, 2013.
- [8] Albert Bifet, Jesse Read, Bernard Pfahringer and Geoff Holmes, "Efficient Data Stream Classification via Probabilistic Adaptive Windows", in Proc. 28th Annual ACM Symposium on Applied Computing, 2013, pp. 801-806.
- [9] Soha Safwat Labib, "A Comparative Study to Classify Big Data Using fuzzy Techniques", in Proc. 5th International Conference on Electronic Devices, Systems and Applications (ICEDSA), 2016.
- [10] Malak El Bakry, Soha Safwat and Osman Hegazy, "A Mapreduce Fuzzy technique of Big Data Classification, in Proc. SAI Computing Conference 2016, pp. 118-128.
- [11] Benjamin Quost and Thierry Denoeux, "Clustering and Classification of fuzzy data using the fuzzy EM algorithm", Fuzzy Sets and Systems, vol. 286, pp. 134-156, 2016.
- [12] Zhenyn Deng, Xiaoshu Zhu, Debo Cheng, Ming Zong and Shichao Zhang, "Efficient kNN classification algorithm for big data", Neurocomputing, vol.195, pp. 143-148, 2016.
- [13] Shichao Zhang, Debo Cheng, Ming Zong and Lianli Gao, "Self-representation nearest neighbour search for classification", Neurocomputing, vol.195, pp. 137-142, 2016
- [14] Ge Song, Justine Rochas, Lea El Beze, Fabrice Huet and Frederic Magoules, "K Nearest Neighbour Joins for Big Data on MapReduce: A Theoretical and Experimental Analysis", IEEE Trans. on Knowledge and Data Engineering, vol. 28, no. 9, pp. 2376-2392, 2016.
- [15] Jesus Maillou, Sergio Ramirez, Isaac Triguero and Francisco Herrera, "kNN-IS: An Iterative Spark-based design of the k-Nearest Neighbours classifier for big data", Knowledge-Based Systems, vol. 117, pp. 3-15, 2017.

- [16] Xindong Wu et. Al., “*Top 10 algorithms in data mining*”, Knowledge and Information Systems, vol. 14, no. 1, pp 137, 2008.
- [17] J. MacQueen, “*Some methods for classification and analysis of multivariate observations*”, in Proc. 5th Berkeley Symp. Math. Stat. Probab., L. M. L. Cam and J. Neyman, Eds. Berkeley, CA: Univ. California Press, 1967, vol. I.
- [18] Junjie Wu, “*Advances in K-Means Clustering*”, Berlin, Heidelberg: Springer-Verlag, 2012.
- [19] Jeffrey Dean, Sanjay Ghemawat , “*MapReduce: A Flexible Data Processing Tool*”, Communications of the ACM, vol. 53 no. 1, pp. 72-77, 2010.
- [20] Morris H. DeGroot and mark J. Schervish, “*Probability and Statistics*”, 4th Edt., Addison-Wesley, 2012.
- [21] UCI Center for Machine Learning and Intelligent Systems, (2017, AUG 01). UC Irvine Machine Learning Repository[Online]. Available: <https://archive.ics.uci.edu/ml/>
- [22] O.L. Mangasarian, W.N. Street and W.H. Wolberg, “*Breast cancer diagnosis and prognosis via linear programming*”, Operations Research, vol. 43, no. 4, pp. 570-577, July-August 1995.
- [23] M. Charytanowicz, J. Niewczas, P. Kulczycki, P.A. Kowalski, S. Lukasik, S. Zak, “*A Complete Gradient Clustering Algorithm for Features Analysis of X-ray Images*”, Information Technologies in Biomedicine, Springer-Verlag, Berlin-Heidelberg, pp. 15-24, 2010.
- [24] F. Alimoglu, E. Alpaydin, “*Methods of Combining Multiple Classifiers Based on Different Representations for Pen-based Handwriting Recognition*”, in Proc. Fifth Turkish Artificial Intelligence and Artificial Neural Networks Symposium (TAINN 96), June 1996.
- [25] James Gosling, Bill Joy, Guy Steele, Gilad Bracha, Alex Buckley, (2017, AUG 01). The Java Language Specification-Java SE 8 Edition [Online]. Available: <https://docs.oracle.com/javase/specs/jls/se8/html/index.html>