An Improved K-Means Clustering Process Solicitation for Mine Blood Donors Information

Anil Audumbar Pise [1,*], Ganesh Shivaji Pise[2], Saurabh Singh [3], Hemachandran K. [4], Shraddhesh Gadilkar[5], Jude Imuede[6], Sandip Shinde[7]

[1]University of the Witwatersrand Johannesburg, South Africa
[2]Assistant Professor in Pune Institute of Computer Technology Pune, India
[3]Assistant Professor, Department of AI and Big data, Woosong University, Daejeon, South Korea
[4]Professor, School of Business, Woxsen University, Hyderabad, India
[5]Associate Engineer, TSYS Global Payments, Pune, India
[6] University of Prince Edward Island, Canada
[7]Department of Computer Engineering, Vishwakarma Institute of Technology, Pune, India

Emails: anipise@deloitte.co.za; ganeshpise143@gmail.com; singh.saurabh@wsu.ac.kr; hemachandran.k@woxsen.edu.in; sgadilkar@tsys.com, jimuede@upei.ca; sandeep.shinde@vit.edu

Abstract
The exponential rise in accidents and the introduction of new, supposedly trendy ways of living have contributed to the dire need for the needy to have an organ or blood transfusion. Circumstances refer to a circumstance where proper care should be taken when collecting the necessary blood or original parts for transfusion, typically in dire circumstances. To determine the distance at which the interested and qualified donors are located, a thorough investigation must be conducted. People are often first categorized according to their blood type, eligibility, and region. Following that, people group together according to locality. A healthy person can safely donate blood twice within 56 days, as this is the minimal time between successful donations that has been established as a norm. The decision to donate an organ is often made after careful consideration of the severity of the situation, the donor's health, and the health of the recipient. Knowledge data finding tasks can be made easier with the help of KEEL, an open-source programme. The graphs that are produced show clearly how the proposed algorithm varies from the standard K-means method. Therefore, it will be quite useful in the present day and could end up saving lots of lives. The necessity to decide ahead of time on the total number of groups is just one of the issues with the K-means clustering method. In practise, it is difficult to anticipate the precise number of clusters. When the number of clusters is small, incongruous clustering is more common, but when the number of clusters is large, like clustering is more common. Thanks to a method called Active Cluster with Modified k-means clustering, which finds the right number of clusters on the fly, the issue is now resolved.

Keywords: Blood Donors; K-means method; Blood transfusion; ACK-means method.

1. Introduction

Numerous real-world industries, like the medical industry, market analysis, advertising, etc., heavily rely on data analysis. There is an urgent necessity today to obtain blood donor information right away because of the high number of accidents and health issues. Since blood bank repositories are expanding quickly, it is getting more and harder to retrieve the necessary data using traditional database techniques [1]. In order to solve this problem, a cluster analysis approach must be used to gather data on blood donors, measure performance using performance analysis metrics, and validate the clustered data using soft computing methods. For the prediction, the target data are selected. By utilizing methods like classification, clustering, association, and regression, data mining is the process of obtaining usable information, patterns, and trends from a huge amount of data [2]. One of the data mining approaches that are helpful for predicting group membership for data instances is classification. A supervised type of machine learning called classification makes use of

Doi : https://doi.org/10.54216/JCIM.130101
Received: April 12, 2023 Revised: August 13, 2023 Accepted: December 04, 2023

previously labelled data. By offering training, we can train the data and forecast its future. Predicting the class to which data may belong is one sort of prediction. Training is carried out using the specified training sample [3].
Clustering or exploratory data analysis is a method of categorization in the absence of labels, or unsupervised classification. Clustering analysis, one of the main data mining analytical approaches, finds the innate classifications in a dataset. Through the process of clustering [4,5], a set of data items is organised into a hierarchy of meaningful subclasses. It is the default distance measure in the k-means clustering technique. Clustering algorithms like K-means, which are based on a partitioning technique, have many practical uses in fields as diverse as economics, image processing, document categorization, and pattern recognition. Early cluster centroids determined using the K-means method can be inaccurate. Since the initial cluster of centroids is formed at random, this precludes the approach from yielding the expected results [6]. The success of the first iteration of the k-means technique hinges on the accuracy of the starting centroid. This work tries to improve the accuracy of the k-means approach by increasing the precision with which the early centroid is calculated [7-8].

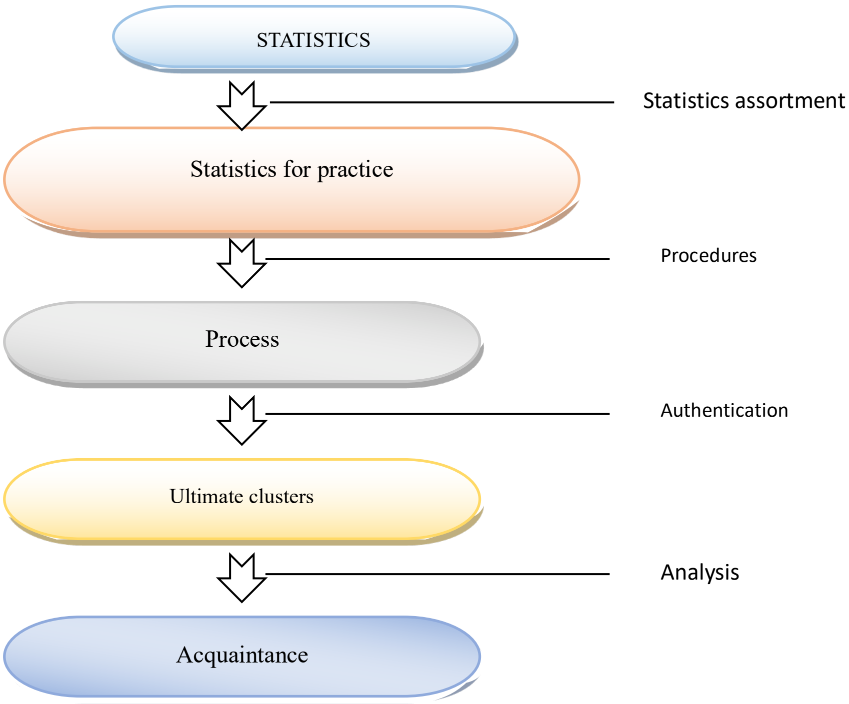

Figure 1: Data mining is the fundamental of KDD Procedure.

Validation of the clusters occurs when it can be determined whether the groups that were formed are dependable or not and whether the data can be correctly identified based on the groupings. The primary three indices, sometimes referred to as testing criteria, that can be used to evaluate each of these is as follows, External indices, internal indices, and relative indices are the first three.
Another data mining method that uses supervised learning and is used to forecast a continuous and numerical target is regression. Regression begins with known values for the data set. It is based on how people learn. By contrasting already known and anticipated values, it calculates the value. Certain model errors, also known as residuals, which are discrepancy between expected and predicted values, can be used to summarise these values [9]. Regression techniques come in two varieties: linear regression and non-linear regression. When a straight line can be utilized to illustrate the relationship between the target and the predictor, linear regression is used. The association in non-linear regression cannot be a straight line. The data can be pre-processed to show this as a linear reaction [10].
To evaluate computational intelligence techniques for DM issues like regression, classification, grouping, and other issues, KEEL is a Java open-source software tool. It includes a wide range of both traditional and modern techniques, such as hybrid models, pre and post-processing procedures, and statistical methodologies for contrasting experiments, among others. Secondly, it enables thorough comparisons between new computational intelligence projects and those that already exist, and KEEL was created for both educational and research objectives [11].
Artificial neural networks are utilized in many scientific and technical applications, and because they roughly resemble complex and non-linear equations, they are also effective tools for electronic analysis [12]. Artificial neural networks' non-linear, non-parametric adaptive learning characteristics make them appropriate for pattern recognition and classification tasks. The system is created using the Neural Networks Tool in MATLAB program to identify the exact

Doi : https://doi.org/10.54216/JCIM.130101
Received: April 12, 2023 Revised: August 13, 2023 Accepted: December 04, 2023

blood group. The workhorse of neural network software is the multilayer feed-forward neural network. This neural network tool can be used to solve problems including pattern recognition, function fitting, and prediction by including a tapped delay line [13]. In this research paper section 2 discussed about related work done and proposed method in subsequent section 3, further result and discussion in section 4 and concluded finally.

2. Related work done

To extract valuable information from the vast amount of data gathered and kept in an organization's databases or repositories, a strong analytical solution is required. Due to this, Knowledge Discovery in Databases (KDD) has emerged. KDD is in charge of turning low-level data into high-level information for decision-making. Data mining is one of the iterative sequence phases of operations that make up knowledge discovery in databases (KDD) [14]. Data mining entails employing numerous algorithmic techniques to discover hidden relationships inside massive data sets. There is a wealth of information stored in healthcare databases, including those at blood banks. The healthcare sector is information-rich but knowledge-poor. That's why [[15]] there's been a shift towards statistical research powered by data in the field of healthcare. Massive amounts of data relevant to healthcare are currently being collected, much as how information is obtained and made available to medical research organisations through the use of computers and other automated methods. Knowledge Discovery in Databases (KDD), which employs data mining methods, has therefore gained popularity among academics in the healthcare business. It allows users to find and capitalise on relationships between seemingly unrelated factors, and to use the data within datasets to create predictions [16].

There are many clustering methods available because there has been a significant amount of research in the area. The studies of many scholars contain thorough analyses of examples of these algorithms [17]. The distance measure being used may affect the outcomes of clustering algorithms. The Euclidean distance is the quantity that clustering algorithms most frequently use. The most well-known clustering algorithm, K-Means, creates non-overlapping clusters. It is believed to be more efficient than hierarchical algorithms. Each cluster has a centroid, which stands for the cluster's overall characteristics [18].

A small number of researchers have worked on the classic K-means clustering algorithm and explored its shortcomings, including estimating the distance between each data object and all groups included in each iteration of the algorithm. In order to solve this problem, they proposed an improved version of the K-means method as well as a more efficient approach to clustering data points. In order to build upgraded k-means and kernelized fuzzy c-means for segmenting brain images, it conveys to numerous local minima that are predicated on highly early clusters enters [19]. They recommended using k-means for the initial cluster entries; however, kernelized fuzzy c-means with the cluster centre initialization strategy fared significantly better than k-means. It is necessary to begin by initializing the number of clusters denoted by k. Subsequently, the initial clusters that are entered are selected at random, and the method is subject to the influence of noise points [20]. In order to solve these problems, they developed an improved version of the k-means algorithm, which consists of a noisy data filter that was developed with the help of a density-based detection approach and noise data characteristics. The cohesiveness of the results of the clustering is enhanced, the influence of noise data on the K-means method is reduced, and processing steps and detection of noise data are added to the technique that was originally used in order to get rid of these noisy data before clustering the cluster. The outcomes of the clustering provide more accurate information [21].

3. The Proposed Method

The main objective of the research work is to evaluate the data mining algorithms' performance analysis and clustering the data of the blood donor using an enhanced k-means method. K-means is a clustering technique used in data mining. Clustering is the process of breaking a dataset into subsets, or groups, wherein items with similar characteristics are placed together. In contrast to classification, clustering relies on unsupervised learning, therefore the final clusters won't be known until after the clustering algorithm has been run. While some clustering algorithms allow the operator to stipulate the desired numeral of clusters, others choose a suitable numeral of clusters automatically.

The K-means technique performs database clustering through an iterative process. The inputs are the numeral of clusters and the means from the beginning, and the output is the means from the end. The cluster means of the beginning and endpoints are given. If the goal of the algorithm is to generate K clusters, then the method will generate K beginning means and K output means.

When K-means clustering is complete, every item in the dataset belongs to exactly one cluster. Finding the cluster with the mean closest to the item is how this cluster is selected. The average of the cluster to which the object under study belongs is taken to be the one with the shortest distance. The K-means method attempts to cluster the data points in a given dataset into the specified number of groups. It iterates until it converges on the best solution to complete the task at hand. The calculated means are modified after each iteration to be more indicative of the final means. The algorithm stops iterating once it reaches convergence. The graphic below depicts the K-means algorithm's predicted convergence. In the

Doi : https://doi.org/10.54216/JCIM.130101
Received: April 12, 2023 Revised: August 13, 2023 Accepted: December 04, 2023

shown case, the algorithm converges in three iterations. The blue dots indicate some initial means that were likely compiled at random. The purple line represents the middle ground.

The K-means clustering algorithm will, in the end, produce red dots that are reflective of the final means. As can be seen in the illustration, the K-means approach iteratively pulls the average closer and closer to the centres of each cluster with each passing iteration. When the approach reaches the points of convergence included within each cluster, it is said to have converged.

K-Means To organize data into meaningful groups, a common unsupervised machine learning approach is clustering [25]. The algorithm divides the data into groups based on their similarities and then tries to find the smallest possible sum of squares for each group.

$$2J(c,\mu)=\sum i=1k\sum j=1n\|xj(i)-\mu i\|2 \qquad (1)$$

In this equation:

- $2J(c,\mu)$ represents a cost function that K-Means tries to minimize. The '2' might be a scaling factor or a part of another formulation. In standard K-Means, this is often just represented as J.
- k is the number of clusters.
- n is the total number of data points.
- $xj(i)$ is the j-th data point in the i-th cluster.
- μi is the centroid of the i-th cluster.
- The summation $\sum i=1k\sum j=1n\|xj(i)-\mu i\|2$ calculates the sum of the squared Euclidean distances of each data point from its respective cluster centroid.

This equation is central to the K-Means clustering algorithm, where the goal is to partition n observations into k clusters in which each observation belongs to the cluster with the nearest mean (centroid), serving as a prototype of the cluster.

The objective function for k-means clustering is to minimize the sum of squared distances between data points and their cluster centroids:

$$\text{Minimize}J=\sum i=1k\sum \mathbf{x}\in Ci\|\mathbf{x}-\boldsymbol{\mu} i\|2 \qquad (2)$$

Where:

k is the number of clusters.

C I represents the data points in cluster i.

μi is the centroid of cluster i

Clustering helps in identifying natural groupings of domains with similar characteristics, which can inform better resource allocation and collaboration strategies. The algorithm iteratively adjusts the positions of the centroids and reassigns data points to clusters to minimize this cost function. K-Means starts by randomly placing data points on a centroid and then iteratively updating those centroid positions. Repeating this process guarantees that the clusters will eventually be uniformly spaced.

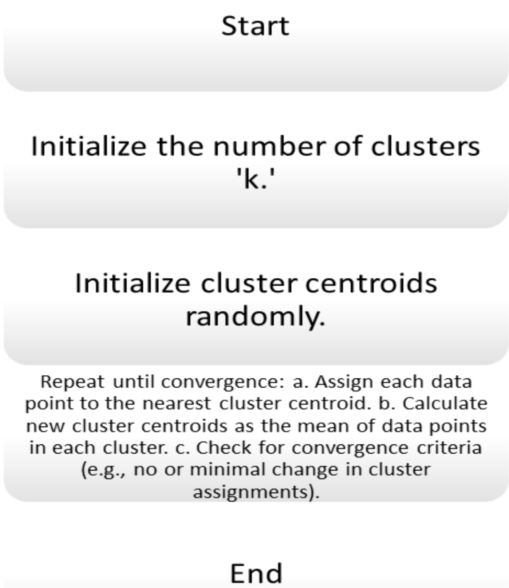

Figure 2: K-Means Cluster Analysis

Doi : https://doi.org/10.54216/JCIM.130101
Received: April 12, 2023 Revised: August 13, 2023 Accepted: December 04, 2023

Figure 2 illustrates unsupervised approach of grouping data into subsets based on their similarities, K-Means clustering is shown in the accompanying flowchart, where cluster centroids are updated repeatedly until convergence is reached.

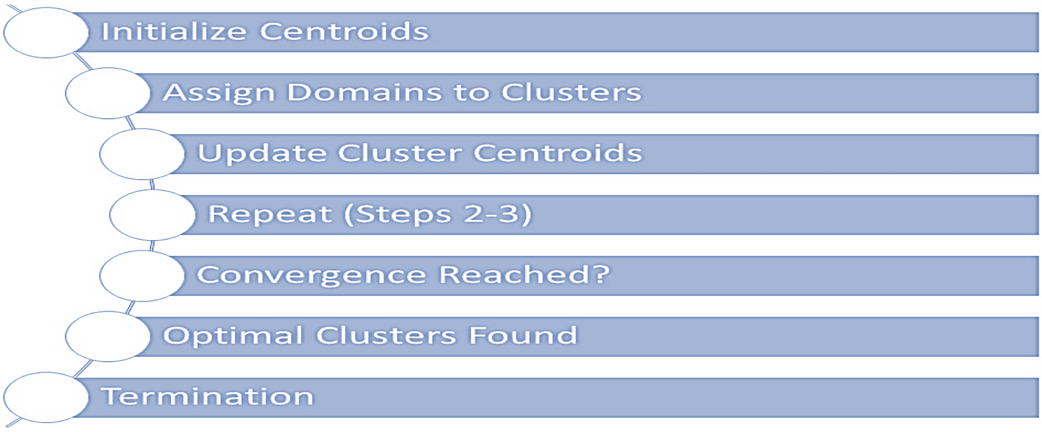

Figure 3: Clustering Algorithms for Domain Grouping.

K-means is a well-liked alternative due to the fact that it is an iterative, numerical, unsupervised, and non-deterministic strategy. Due to the ease of use and quickness of the method, it has been demonstrated to be an extremely efficient approach that can produce satisfactory clustering results in a variety of real-world situations.
The result is a probability distribution, denoted by the letter P, over the various tumor classifications.

1. Feature maps: $Fk=I\otimes Kk+Bk$ F (3)
2. Where Fk represents the k-th feature map, K_k is the k-th convolution kernel, and B_k is the bias.
3. Pooling: $Pk=FkS$ Where Pk represents the pooled feature map, and S is the pooling size.
4. Flattening: Fi=Flatten (Pi) (4)
5. Dense Layer: $\sigma(Wi\cdot Fi+bi)$ (5)

   Where σ is the activation function, W_i is the weight matrix, and b_i is the bias for the i-th dense layer.
6. Output: $P(Ci)$=Softmax(Di) (6)

   Softmax transforms dense layer outputs into class probabilities.

Due to the need to organise the ever-increasing amounts of data and derive actionable insights from them, clustering techniques are used in a wide variety of fields. Some of these fields include artificial intelligence (AI), biology, customer relationship management (CRM), data compression, data mining, information retrieval, image processing, machine learning, marketing, medicine, pattern recognition, psychology, statistics, and many more.

4.1.Algorithm for k-means clustering:

D = d1, d2,...,dn, k //a collection of n data points to be processed. k Cluster size requirement

What we get out of this is a collection of k clusters.

Start by:

Select k random statistics points from dataset D to serve as initial centres;

2. Iterate

2.1 Place di in the cluster whose centroid is closest to di;

Until the convergence requirement is reached, step 2.2 is to conclude the new cluster mean.

End

Doi : https://doi.org/10.54216/JCIM.130101
Received: April 12, 2023 Revised: August 13, 2023 Accepted: December 04, 2023

K-means algorithm. Unlike the K-means clustering algorithm, initial centroids are not chosen at random. The remaining technique is identical to the well-known K-means method of cluster analysis. Specifically, this is used in the blood group dataset so that we may preserve the pre-existing data. When a large volume of blood is needed quickly, it is collected in clusters based on the victim's location and blood type. In this method, the enhanced clustering algorithm processes values from many dimensions. Each data value di may have several columns of information labelled di1, di2,..., dim, where m is the total number of characteristics. First, we find the column that contains the largest range, where range is the difference between the largest and smallest element in the column.

To enhance the quality of the clusters we have proposed an Active Cluster using a modified k-means clustering technique. Based on the current cluster, we make adjustments to the k-means algorithm. When thinking about real-world applications, knowing the exact number of clusters at runtime is crucial.

4.2. Algorithm for ACMk-Means clustering

1. Select k items at random from set D to serve as seed nodes for the clusters.
2. Repeat
3. (Re)assign each item to the cluster to which it most closely belongs based on the average value.
4. Find the average of all the numbers and use that as the new mean.
5. as long as there is still movement.
6. Proceed to Step 2 if the number of clusters is unchanging.
7. Find the distance between clusters.
8. Quantify the gap between clusters.
9. If the previous intra-cluster distance was greater than the current one, and the previous inter-cluster distance was smaller than the current one, then proceed to step 10.
10. K=k+1 Proceed to the first step.
11. Stop.

Distances between and within clusters are computed using this approach. It is reasonable to assume that a cluster is good if its intra-cluster distance is small. Cluster quality can be demonstrated to be low if the need arises to "fix" the cluster count. The number of clusters in the ACMK-means method is dynamic and is based on the quality of the cluster output during runtime. The k-means and ACMk-means methods were used to evaluate the quality of the clusters in this paper. The clusters in an ACMk-means method can be fixed or a minimum number of clusters can be specified in the input. By adding one to the cluster number at each iteration, this algorithm generates new clusters until the cluster quality meets the threshold.

5. Result and Analysis

Here, we discuss the steps we took to put our system into action. The proposed method actually works by just opening the file and reading in the dataset. The procedure is implemented based on the participant's Blood Group and selected location. Since the clusters are produced depending on both geography and blood group, we can use it to identify those persons who live in closer proximity to the receptor. Different dataset records are used to evaluate the effectiveness and precision of the original K-means and the Enhanced K-means clustering algorithms. The sizes of data sets might vary greatly. K-means and Enhanced K-means are both fed the identical sets of data and compared against one another. Both methods need inputs such as the total number of clusters, the type of blood needed, and the location. In the original k-means algorithm, the initial centroids are chosen at random; in the Enhanced k-means algorithm, they are computed using meaningful formulas.

Clusters obtained from trials are compared to a dataset of donors that meet certain criteria (blood type, geographic region, etc.) to see how well they cluster. Each experiment's success rate and duration are calculated and tabulated.

Doi : https://doi.org/10.54216/JCIM.130101
Received: April 12, 2023 Revised: August 13, 2023 Accepted: December 04, 2023

Table I: Enactment assessment of the two algorithms for 300 dataset proceedings.

	Record Dataset	Cluster count	Inter cluster	Intra cluster
k-means	300	3	0.063	0.023
ACMk-means	300	5	0.075	0.016

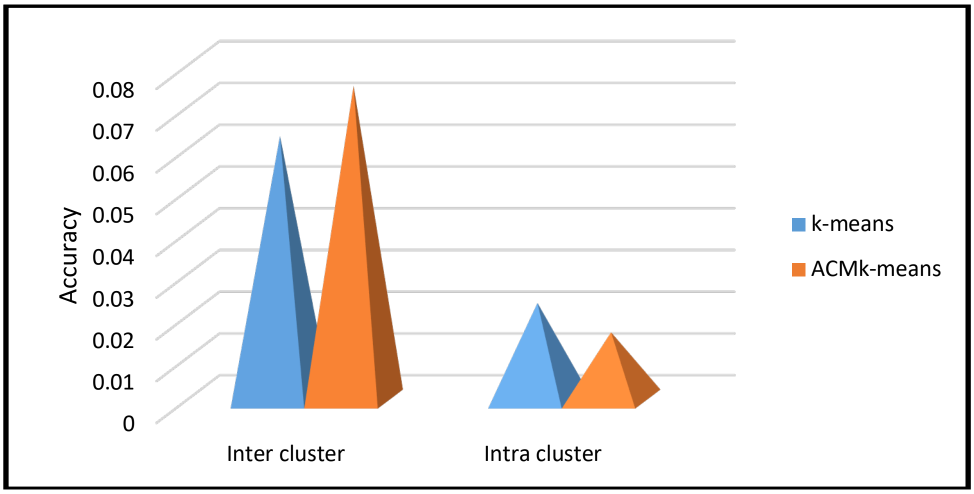

Figure 4:Performance comparison of the two algorithms for 300 dataset records.

Both table 1 and picture 4 illustrate how the K means technique and the suggested algorithms compare to one another in terms of the amount of time required to analyse datasets containing a greater number of records. We observed that the amount of time required to construct clusters became increasingly more extensive as the size of the dataset rose. When only the most essential dataset records are used, the Enhanced K-means approach only requires a small amount of the total available time to create a cluster. The performance of the new algorithm is superior to that of the K-means approach as it was originally implemented. The proposed technology has the potential to rapidly establish clusters, which would ultimately save a great number of lives in the real world. If we locate initial centroids in a meaningful way that enhances the performance of the system that helps to identify available donors with the required blood group in the nearby place, then we will be able to serve a greater number of donors to donate blood in a shorter length of time. This will allow us to meet the needs of a greater number of people who need to give blood.

Table 2: Enactment assessment of the two algorithms for 500 dataset proceedings.

	Record Dataset	Cluster count	Inter cluster	Intra cluster
k-means	500	5	0.034	0.014
ACMk-means	500	6	0.066	0.012

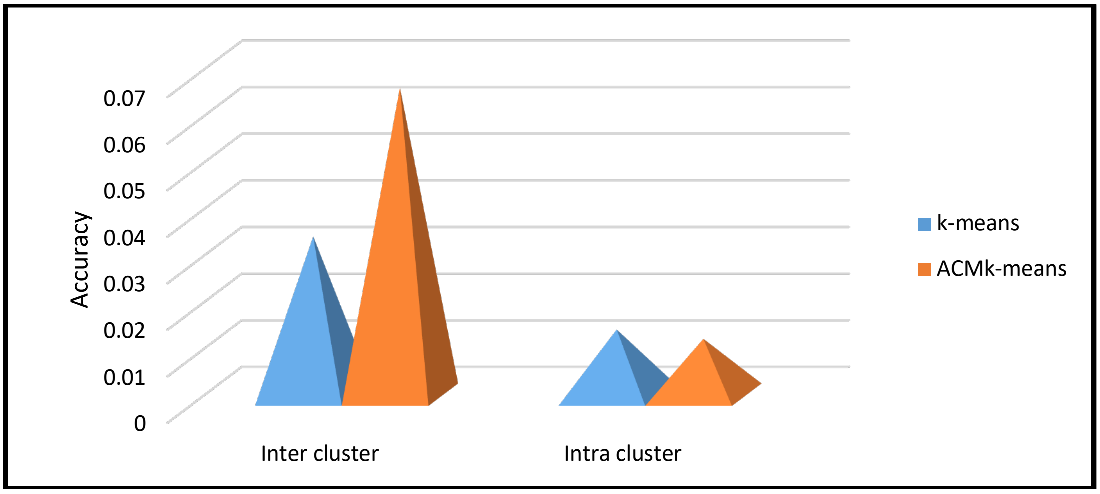

Figure 5:Performance comparison of the two algorithms for 500 dataset records.

Table 3: Enactment assessment of the two algorithms for 700 dataset proceedings.

	Record Dataset	Cluster count	Inter cluster	Intra cluster

Doi : https://doi.org/10.54216/JCIM.130101
Received: April 12, 2023 Revised: August 13, 2023 Accepted: December 04, 2023

k-means	700	7	0.021	0.005
ACMk-means	700	8	0.023	0.006

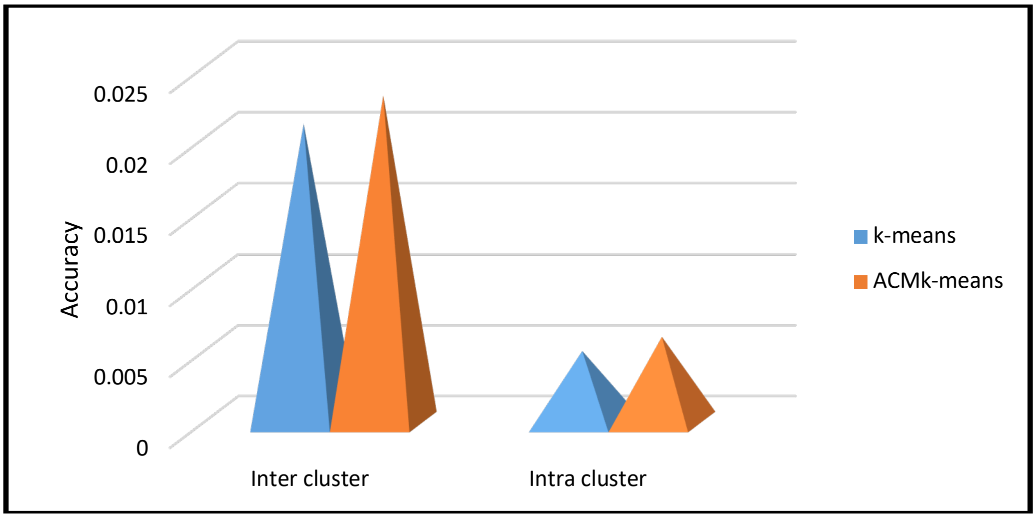

Figure 6:Performance comparison of the two algorithms for 700 dataset records.

The data used in the experiment are completely made up. For this check, we'll use records 300, 500, and 700 from the data set at random. There were 300 records in the data set displayed in table 1. The findings demonstrate the computational efficacy of the Active Cluster with Modified K-means method for evaluating cluster quality. The ratio of inter- to intra-cluster distance should determine a cluster's quality. The internal minimum that must represent a concentration of excellence. When we increase the number of records in the dataset and the number of clusters, the intra cluster comparison between the original k-means and the new k-means becomes more pronounced, as shown in tables 1, 2, and 3. The Active Cluster with Modified KACMK-means method provides the best possible performance and clustering density.

In terms of cluster quality, the Active Cluster with Modified K-means method performs better than the more Means technique. Inter-cluster and intra-cluster distances should be used to determine the quality of a cluster. The internal minimum that must represent a concentration of excellence. We imply and The optimal number of clusters is determined by the Active Cluster with Modified K-means algorithm, which also provides superior efficiency. However, with an unknown data set, it begins with the minimum number of clusters supplied by the user and then iteratively increases the number of clusters after each round of computations.

6. Conclusion

However, picking the right clustering algorithm isn't always easy. It's challenging even to narrow down to the best possible options for closing off a certain dataset. The majority of algorithms make the assumption of a pre-existing, underlying structure in the data. The nature of the data and the cluster that is being sought after are two crucial factors. The nature of the data and the resources needed by the algorithm should also be taken into account. A new blood group and geographically-based K-means clustering algorithm is proposed in this report. If you need to find a list of people who share a specific blood type and where they live, the Enhanced K-means method is a great tool to employ. Some of the temporal constraints that plague the popular K-means clustering approach are mitigated here. The resulting graphs make it easy to see how the suggested algorithm differs from traditional K-means. As a result, it will be extremely helpful in the present day, potentially saving many lives. The K-means clustering technique has a number of problems, one of which is the need to predetermine the total number of groups. It is challenging to determine the exact number of clusters in advance in practice. There will be more instances of incongruous clustering if the number of clusters is low, and more instances of like clustering will occur if the number of clusters is high. Because of this, the problem has been fixed thanks to an approach known as Active Cluster with Modified k-means clustering, which determines the appropriate total number of clusters dynamically.

References

[1] Uma Boregowda, Venugopal R Chakravarthy, “A Hybrid Task Scheduler for DAG Applications on A Cluster of Processors”, 978-1-4799-4363-0/14 2014 IEEE

[2]. TuLinLi, Deng YanNi, Chu SiYong, “A K-means Clustering Algorithm Based on Double Attributes of Objects”, 978-1-4673-7143-8/15 2015 IEEE

[3]. David P. Pancho, Jose M. Alonso, Jes´usAlcal´a- Fdez “A new fingram-based software tool for visual representation and analysis of fuzzy association rules”, ,10.1109/FUZZ-IEEE.2013.6622478 July 2013 IEEE

Doi : https://doi.org/10.54216/JCIM.130101
Received: April 12, 2023 Revised: August 13, 2023 Accepted: December 04, 2023

[4]. Fuyuan Cao, Jiye Liang and Liang Bai “A new initialization method for categorical data clustering”, 2009 Elsevier Ltd.
[5]. Muamer N. Mohammada, NorrozilaSulaimana, Osama AbdulkarimMuhsinb “A Novel Intrusion Detection System by using Intelligent Data Mining in WEKA Environment”, , 1877-0509 2010 Published by Elsevier Ltd. doi:10.1016/j.procs.2010.12.198
[6]. Aris- Kyriakos Koliopoulos, ParaskevasYiapanis, FiratTekiner, Goran Nenadic, John Keane “A Parallel Distributed Weka Framework for Big Data Mining using Spark”, 2015 IEEE International Congress on Big .
[7] D. C. McElfresh, C. Kroer, S. Pupyrev, E. Sodomka, K. Sankararaman, Z. Chauvin, N. Dexter, and J. P. Dickerson, "Matching Algorithms for Blood Donation," arXiv preprint arXiv:2108.04862, Aug. 2021

[8] B. Venkateswarlu and G.S.V.Prasad Raju “Mine Blood Donors Information through Improved K-Means Clustering”, International Journal of Computational Science and Information Technology (IJCSITY) Vol.1, No.3, 2320-8457, August 2013
[9] B. Venkateswarlu and G.S.V.Prasad Raju “Organ Donor Identification through Improved K- Medoids Clustering”, International journal of computer science and technology, volume 5, issue 3, 2229-4333, July – Sept 2014.
[10] Sarker IH (2022) Ai-based modeling: Techniques, applications and research issues towards automation, intelligent and smart systems. SN Computer Science 3(2):1–20
[11] A. Erlansari, R. Effendi, F. F. C., A. Wijanarko, B. Susilo, and R. Hardiansyah, "Backpropagation and Fuzzy Algorithm Modelling to Resolve Blood Supply Chain Issues in the COVID-19 Pandemic," arXiv preprint arXiv:2109.02645, Sep. 2021
[12] Sarker IH (2021) Data science and analytics: an overview from data-driven smart computing, decision-making and applications perspective. SN Computer Science 2(5):1–22
[13] Sarker IH, Colman A, Han J, Watters PA (2021) Context-aware machine learning and mobile data analytics: automated rule-based services with intelligent decision-making. Springer Nature, Switzerland
[14] V. Roy and S. Shukla, "A NLMS Based Approach for Artifacts Removal in Multichannel EEG Signals with ICA and Double Density Wavelet Transform," 2015 Fifth International Conference on Communication Systems and Network Technologies, 2015, pp. 461-466, doi: 10.1109/CSNT.2015.61.
[15] Rahim MdS, Ahmed T (2017) An initial centroid selection method based on radial and angular coordinates for
k-means algorithm. In 2017 20th International Conference of Computer and Information Technology (ICCIT), 1–6. IEEE
[16] Kumar A, Gupta SC (2015) A new initial centroid finding method based on dissimilarity tree for k-means algorithm. arXiv preprint arXiv:1509.03200
[17] Mahmud MdS, Rahman MdM, Akhtar MdN (2012) Improvement of k-means clustering algorithm with better initial centroids based on weighted average. In 2012 7th International Conference on Electrical and Computer Engineering, 647–650. IEEE
[18] S. K. Dash, B. K. Tripathy, and A. Acharya, "An Improved K-Means Clustering Process Solicitation for Mine Blood Donors Information," International Journal of Computer Applications, vol. 176, no. 34, pp. 1–6, Jun. 2020
[19] Lakshmi MA, Daniel GV, Rao DS (2019) Initial centroids for k-means using nearest neighbors and feature means. In Wang J, Reddy GRM, Prasad VK, Reddy VS (eds), Soft Computing and Signal Processing, 27–34, Singapore. Springer Singapore
[20] Sawant KB (2015) Efficient determination of clusters in k-mean algorithm using neighborhood distance. The International Journal of Emerging Engineering Research and Technology 3(1):22–27
[21] Motwani M, Arora N, Gupta A (2019) A study on initial centroids selection for partitional clustering algorithms. In Hoda MN, Chauhan N, Quadri SMK, Srivastava PR (eds), Software Engineering, pages 211–220, Singapore. Springer Singapore.

Doi : https://doi.org/10.54216/JCIM.130101
Received: April 12, 2023 Revised: August 13, 2023 Accepted: December 04, 2023