RESEARCH ARTICLE

European Journal of Information Technologies and Computer Science
www.ej-compute.org

A Hybrid Clustering Model for Determining Financially Needy Students for Sponsorship

Samuel King Opoku and Asare Yaw Obeng

ABSTRACT

Many organizations in the world are seeking effective means to provide financial assistance to brilliant but needy students. Many research works have been conducted to determine needy students based on classification methods which fail to separate financially stable students from needy ones as the class labels for the classification methods are affected by manual reviews, randomness and the discretion of the selection committee. This paper presents a hybrid clustering method for determining needy students based on income and expenditure using the K-Means and Expectation-Maximization methods with the logical AND operator to eliminate the limitations associated with the classification methods. The study reveals that hybridizing clustering methods yield better results. Also, for a normalized data categorization, the Euclidean distance gives better results than the Manhattan distance in the K-Means algorithm.

Keywords: Clustering, Euclidean distance, Expectation-Maximization, K-Means, Manhattan distance, needy students, sponsorship.

Published Online: July 4, 2023

ISSN: 2736-5492

DOI :10.24018/ejcompute.2023.3.3.106

S. K. Opoku*
Computer Science Department, Kumasi Technical University, Ghana.
(e-mail: samuel.kopoku@kstu.edu.gh)

A. Y. Obeng
Computer Science Department, Kumasi Technical University, Ghana.
(e-mail: yaw.aobeng@kstu.edu.gh)

Corresponding Author

I. Introduction

Many organizations in the world are seeking effective means to provide financial support to brilliant but needy students. The normal process of selecting students to sponsor usually requires a manual review of the submitted application forms, the discretion of the selection committee and randomness. The basic parameter for identifying needy students in many countries is the family economic survey [1]. However, different parameters are chosen by different countries and different regions. The United States identifies needy students by using family income. Regarding Japan, income and assets are used to determine the economic status of families. Ugandans depend on the father's job and the number of vehicles or the type of vehicle to determine a family's economic status. Nigerians depend on family occupation and family size to determine the economic status of the various families. This underlining problem is the implementation of a generalized framework for determining needy students.

Previous works for determining needy students are based on classification methods such as factorization machines [1], deep learning [2], decision tree C5.0 algorithm [3], random forest and logistic regression [4]. These related works determined needy students from family income, family expenditure and level of borrowing. Students' lifestyle in prioritizing study activities [5] also has a part to play in managing students' finances. The above models required class labels in a supervised learning environment [6], [7] to determine needy students. These models, however, require the existing methods for determining needy students which are affected by manual review, randomness and discretion of a selection committee. This paper determines needy students based on income generation and expenditure on critically unavoidable circumstances through a hybrid clustering mechanism.

Clustering is the process by which systems or models intelligently group similar objects without class labels or under unsupervised machine learning mechanisms. There have been numerous categorizations of clustering mechanisms over the years. The work of [8] classified clustering mechanisms into four groups. Exclusive clustering (or hard clustering) ensures that each datum belongs to a definite cluster. Overlapping clustering (or soft clustering) categorizes a datum into two or more clusters using membership degree. The hierarchical clustering mechanism starts by making every datum as a cluster as a bottom-up mechanism (agglomerative) or starts with the entire data set as a single cluster and starts splitting into various sub-clusters as a top-down mechanism (divisive). Probabilistic clustering assigns a subjective probability for a datum to belong to a membership of a cluster.

Reference [9], however, categorized clustering mechanisms into five groups. The partitioning clustering splits data points into k distinct partitions representing k clusters based on such functions as minimizing the square error. The K-Means algorithm is an example of a partitioning clustering mechanism. The density-based clustering finds clusters in regions with higher concentrations of data whilst considering spare regions as noise or outliers. The grid-based clustering mechanisms use grid data structures to quantize space into a finite amount of grids as clusters. The model-based clustering methods locate clusters by clustering the density function based on standard statistics whilst taking care of outliers. Reference [10] added text-based and soft computing clustering methods to some of the above categorizations to obtain eight groups of clustering methods. Soft computing clustering uses the evolution approach, fuzzy approach and simulated annealing approach to cluster data.

Different researchers have combined different categories [11], [12] whilst others have provided variations to the existing mechanisms in their work [13]. This study selected an exclusive (or hard), non-hierarchical, partition-based clustering mechanism. Hierarchical clustering lacks interpretability about the clusters obtained and inability to make corrections once the splitting or merging has been done [10], thus, it was ignored. Since the data to be used in this work has been cleaned without noise and every data is required to be categorized, the density-based method was ignored. Grid-based algorithms are usually density-based [9] and result in low accuracy levels as they inefficiently handle boundary point problems [14].

Data in this study should exclusively be categorized into two groups, either needy students or not needy (financially stable students). Because of this, overlapping clustering mechanisms were ignored in this study. The Expectation-Maximization (EM) mechanism which is a Gaussian Mixture Model can be implemented as a hard clustering method [15]. Thus, K- Means and hard EM algorithms were chosen for this study. The results from both algorithms were hybridized using the logical AND operator. The remaining parts of this paper are divided into three sections. Section II looks at the methodology of the study. Section III discusses the results and evaluates the method used in this study and Section IV concludes the paper

II. Materials and Methods

A. Data Collection

Data was collected from 1860 respondents. Interestingly, all the respondents claimed that they were needy and required sponsorship to complete their university education. These students have varying conditions. The questionnaire demanded answers to the following questions:

- Number of parents taking care of you
- Income of your parents
- Are you working whilst schooling?
- Your income as a student
- Debt amount of your parents
- Number of your household having medical problems
- Cost of medical expenses per day
- Number of siblings in education
- Cost of siblings in education

It can be seen that the data collected is based on income generation and expenditure. The first four responses look at income generation whilst the remaining responses look at how the generated income is spent on critically unavoidable issues.

B. Data Processing and Normalization Mechanism

Since all the respondents claimed that they were needy and the study did not believe same, the class label was ignored. Rejecting the class label, forced the problem to be categorized or clustered. The data submitted by the respondents was normalized. Normalizing data ensures fast convergence with an increased level of accuracy [16], [17].

The benchmark for setting up the ranges of data was based on the economic status of the family. The international poverty line for determining the economic status of a population is set to USD 1.90 [18]. However, the population with USD 4.00 are the "floating class" [19] and they can slide back to the international poverty line. The secured middle-class population earns income between USD 4.00 and USD 10.00 [19], [20] and they can handle their financial burdens and obligations successfully.

The number of parents taking care of the student ($P \in \mathbb{Z}^+, P \leq 2$) was normalized as 0, 1 and 2 where 0 represents a student without a guardian. Such students are orphans. 1 represents students with single parents and 2 represents students with both parents. A parent in this study also represents a guardian.

The income of the parent ($IP \in \mathbb{R}^+, IP \geq 0$) was categorized into ranges and was normalized as 0, 1 and 2. Irrespective of the number of parents, the income range falls under one of the three categories. Zero (0) indicates that the total income of the parent(s) is less than USD 4.00, 1 indicates that the income of the parent(s) is between USD 4.00 and USD 10.00 a day and 2 indicates that the income of the parent(s) is greater than USD 10.00 a day.

The third question regarding whether the student is working whilst schooling has two nominal answers, namely, yes and no. The study normalized the response as 0 for no and 1 for yes. Like the income range of parents, the income range of students ($IS \in \mathbb{R}^+, IS \geq 0$) was normalized with three values as either 0, 1 or 2. Zero (0) indicates that the student is not working or s/he is earning less than USD 4.00 a day. One (1) indicates that the student is working and s/he is earning between USD 4.00 and USD 10.00 a day. Two (2) indicates that the student is working and s/he is earning more than USD 10.00 a day.

The debt amount of the parents ($PD \in \mathbb{R}^+, PD \geq 0$) was also normalized with three values as 0, 1 and 2. 0 indicates that the parents are not in debt or they are in debt such that the total amount to be paid daily is less than USD 4.00. 1 indicates that the parents are in debt with daily payments between USD 4.00 and USD 10.00. 2 indicates that the debt of the parents requires a daily payment of more than USD 10.00.

The next question demanded the number of household members with chronic medical conditions requiring constant medication ($M \in \mathbb{Z}^+,\ M \geq 0$). This was not normalized. However, the daily medical cost ($MC \in \mathbb{R}^+,\ MC \geq 0$) was normalized using 0, 1 and 2 where 0 indicates a daily cost of less than USD 4.00, 1 indicates a daily cost between USD 4.00 and USD 10.00 and 2 indicates a daily cost of more than USD 10.00

The number of siblings ($S \in \mathbb{Z}^+,\ S \geq 0$) being taken care of in education was also not normalized. However, the daily cost of educating the siblings ($EC \in \mathbb{R}^+,\ EC \geq 0$) were normalized as 0, 1 and 2 where 0 indicates a daily cost of less than USD 4.00, 1 indicates a daily cost between USD 4.00 and USD 10.00 and 2 indicates a daily cost of more than USD 10.00

Certain attributes of the data collected were ignored. Some were irrelevant to the clustering problem, though they were used to validate the information provided by the respondents. Attributes that could also be inferred from others were also ignored. The gender and age data of the questionnaire were considered irrelevant to the clustering problem and they were therefore ignored. The number of parents was ignored since its consequential effect is found on the income of the parents. The income of the student was enough to handle the question of whether the student was working whilst schooling. Thus, whether the student was working whilst schooling was ignored. Similarly, the number of households with medical conditions and the number of siblings in education were both ignored as their implications were found in medical and educational costs respectively. Thus the resulting attributes selected for the clustering problem were:

- Income of parent(s)
- Income of student
- Debt of parent(s)
- Cost of medical condition
- Cost of educating siblings

The remaining data set created records with duplications after the normalization process which required further cleaning by removing the duplicate records whilst ensuring that there was no missing value in the records. The data set was reduced to 243 unique records which were used in the clustering process.

C. *Working Principles of the Hybrid Clustering Method*

The study depended on two clustering methods for categorizing the data into financially stable students and financially handicapped students. The methods are K-Means and Expectation-Maximization (EM).

The most crucial aspect of the K-Means algorithm is the determination of similarity or dissimilarity among the data points. This is usually achieved by computing the distance between the chosen centroids and the data points. Distance computation has been useful in many applications both in machine learning and non-machine learning applications [21]. Given that R and Q are two points in n-dimensional space, the Minkowski distance, M, which is the generalized formula for distance computation is given as:

$$M = \left(\sum_{i=1}^{n} |R_i - Q_i|^p\right)^{\frac{1}{p}}$$

The value of p determines the type of distance formula used. There are three common values of p used.

When p = 1, we have the Manhattan Distance as

$$Manhattan\ Distance = \sum_{i=1}^{n} |R_i - Q_i|$$

When p=2, we have the Euclidean Distance as

$$Euclidean\ Distance = \left(\sum_{i=1}^{n} (R_i - Q_i)^2\right)^{\frac{1}{2}}$$

When p = ∞, we have the Chebyshev Distance which gives

$$Chebyshev\ Distance = Max\ (|R_i - Q_i|)$$

The following steps are required in K-Means implementation:

Begin

1. Determine the number of centroids, say k.
2. Find the best values for the centroids.
3. Pick a data point
4. Calculate the distance between the centroids and the data point.
5. Assign the data point to the cluster with the least distance computation between the data point and the centroids.
6. If the least distance computation between the data point and two more centroids then
   6.1 Place the data point in any of the clusters
7. End if
8. Pick the next data point
9. If there is more data point
   7.1 Go to step 4
10. End if
11. If the assignment in steps 5 through 7 was successful or occurred then
    11.1 Go to step 14
12. Else
    12.1 Go to step 16
13. End if
14. Calculate the variance and place a new centroid for each cluster
15. Go to step 3
16. Finish, the model is ready

End

The Chebyshev distance does not consider all the elements in the n-dimensional space and therefore, it was ignored. The Manhattan distance is longer than the Euclidean distance. For instance, let us consider two points R and Q in 2-dimensional space, the Euclidean distance is $\sqrt{(\Delta R^2 + \Delta Q^2)}$ whereas the Manhattan distance is $|\Delta R| + |\Delta Q|$. Squaring both equations, the square of the Euclidean distance becomes $\Delta R^2 + \Delta Q^2$ whereas the square of the Manhattan distance becomes $(|\Delta R| + |\Delta Q|)^2$. Expanding the square of the Manhattan distance shows that it is greater than the square of the Euclidean distance such that $\Delta R^2 + \Delta Q^2 + 2|\Delta R|.|\Delta Q| > \Delta R^2 + \Delta Q^2$ due to the $2|\Delta R|.|\Delta Q|$ factor. The normalization process of the data reduces the marginal values of the elements in the n-dimensional space which also affect the distance computation. The Manhattan distance computation will not give a befitting set of clusters. The Euclidean distance which turns to compute the shortest path between the data points and the centroid will be appropriate for this study. However, clustering the dataset with Manhattan distance will be used as a control experiment.

The Expectation-maximization (EM) algorithm combines other unsupervised machine learning algorithms with two stages. The first stage estimates the missing variables. This stage is the expectation or estimation stage. The second stage which is the maximization stage optimizes the parameters of the model using the maximum likelihood estimation (MLE) method that achieves a value that fits the data. The processes are repeated until the convergence of the values occurs or is less than the tolerance error.

Given the dataset for the chosen attributes of the study $D_{i,j}$ such that $\forall i,\ i \in \mathbb{Z}^+, i \leq 5$, the five chosen attributes and $\forall j,\ j \in \mathbb{Z}^+, j \leq 243$, the dataset instance numbers. Also, $D_{i,j} \in \{0, 1, 2\}$, the normalized values for specific attribute i and dataset instance number j. Let $C_j \in \{0, 1\}, \forall j \in \mathbb{Z}^+, j \leq$

European Journal of Information Technologies and Computer Science
www.ej-compute.org

243, be the cluster labels for each dataset instance number j such that $C'_j s$ are unknown or unobserved.
The data now turns

$$\begin{matrix} D_{1,1} & D_{1,2} & \dots & D_{1,5} & C_1 \\ D_{2,1} & D_{2,2} & \dots & D_{2,5} & C_2 \\ & & \vdots & & \\ D_{243,1} & D_{243,2} & \dots & D_{243,5} & C_{243} \end{matrix}$$

The dataset consists of the known $D_{i,j}'s$ as D only is incomplete. Let S = (D, C) be a complete data distribution such that S depends on a parameter α such that L(α) denotes the likelihood function, then the EM algorithm seeks to find MLE for the parameter α which determines the $C'_j s$.

The following steps are required in EM implementation:
Begin
1. Initialize the model with an incomplete dataset or variables as L (α | D, C) where C is the incomplete variable and α is the missing parameter.
2. Let t=0 be an initial estimate for α as $\tilde{\alpha}^t$
3. Guess the incomplete variable from the observed data. Thus, given D and $\tilde{\alpha}^t$, determine the conditional density f(c | D, $\tilde{\alpha}^t$) for the completion variables and calculate the expected log-likelihood value by
   $$Q(\alpha \mid \tilde{\alpha}^t) = \int \ln(f(D, c \mid \alpha)) \times f\left(c \,\middle|\, D, \tilde{\alpha}^t\right) dy$$
4. Use the completion variables to update the parameter of the model as $\tilde{\alpha}^{t+1}$
5. If missing variables converge or $\|\tilde{\alpha}^{t+1} - \tilde{\alpha}^t\| \leq$ error tolerance then
   5.1 Model development is done
6. Else
   6.1 t = t + 1
   6.2 Go to step 3
7. End if

End

Both K-Means and EM algorithms have limitations but they did not affect the study. The K-Means algorithm is limited by the determination of the initial number of centroids. This study determined the number of centroids before the start of the model development. The EM model, though uses both forward and backward probabilities and converges slowly, it is often guaranteed that the likelihood value is enhanced after each iteration.

Two clusters are generated by both K-Means and EM algorithms. Cluster 0 (or simply 0) represents needy students whereas cluster 1 (or simply 1) represents not needy students (other words, financially stable students). The final categorizations are obtained using the logical AND operator such that $\forall D_{i,j} \exists C_j^{K-Means}$ and C_j^{EM} such that $C_j^{K-Means}$, $C_j^{EM} \in \{0, 1\}$ where $C_j^{K-Means}$ is the cluster generated by the K-Means algorithm and C_j^{EM} is the cluster generated by the EM algorithm for the dataset $D_{i,j}$. If $\forall j \in \mathbb{Z}^+, j \leq 243$, $C_j^{K-Means} = C_j^{EM} = q \in \{0, 1\}$ then $D_{i,j} \in q$. However, if $\forall j \in \mathbb{Z}^+, j \leq 243$, $C_j^{K-Means} \neq C_j^{EM}$ then $D_{i,j} \in cluster\ 1$ which indicates that it is false that the student is needy.

The WEKA software was used in this study due to its flexibility and user-friendly interface. The study seeks to work with the categorized data and it is easy to obtain the cluster label assigned to each data instance with additional coding. WEKA does not require any programming skills to use. The major disadvantage of WEKA software has to do with the size of the dataset. Fortunately, this limitation did not affect this study as the dataset was normalized and the size was reduced to 3 KB.

III. Results and Evaluation

Given the initial centroids values for Cluster 0: 0, 0. 0, 0, 0 and Cluster 1: 2, 2, 2, 0. 0, the final clusters produced the following results for the K-Means algorithm using the Euclidean distance in Table I.

TABLE I: Centroid Values of K-Means Algorithm after Clustering using Euclidean Distance

Attributes	Cluster 0	Cluster 1
Parents Income	0.5	1.3007
Student Income	0.3	1.4118
Parent Debt	1.2	1.4118
Medical Cost	1	1
Siblings Education Cost	1	1

Out of the 243 datasets, 90 of them representing 37% of the total dataset were categorized as cluster 0 (needy students). On the other hand, 153 of the total datasets representing 63% were categorized as cluster 1 (financially stable students). Fig. 1 shows a graphical representation of the clusters when the instance number is plotted against the instance number

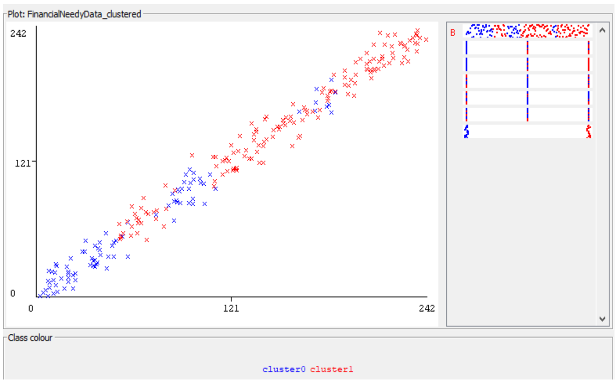

Fig. 1. K-Means Clustered Dataset.

From Fig 1, the dataset coloured blue represents elements of Cluster 0 whereas the dataset coloured red represents elements of Cluster 1. The sum of squares within-cluster distance after clustering was 162.0.

Table II shows the metrics of Cluster 0 and Cluster 1 in terms of the means and standard deviations of the attributes when the EM algorithm was used for clustering.

Out of the 243 datasets, 81 of them representing 33% of the total dataset were categorized as Cluster 0 (needy students). On the other hand, 162 of the total datasets representing 67% were categorized as Cluster 1 (financially stable students). Fig 2 shows a graphical representation of the clusters when the instance number is plotted against the instance number.

TABLE II: CLUSTERING METRIC VALUES OF EM ALGORITHM

Attributes	Metrics	Cluster 0	Cluster 1
Parents Income	Mean	0.9999	1.0062
	Stand. Dev	0.8165	0.8202
Student Income	Mean	0.0005	1.4914
	Stand. Dev	0.0228	0.5114
Parent Debt	Mean	1.3333	1.3333
	Stand. Dev	0.6667	0.6667
Medical Cost	Mean	1	1
	Stand. Dev	0.8165	0.8165
Siblings Education Cost	Mean	1	1
	Stand. Dev	0.8165	0.8165

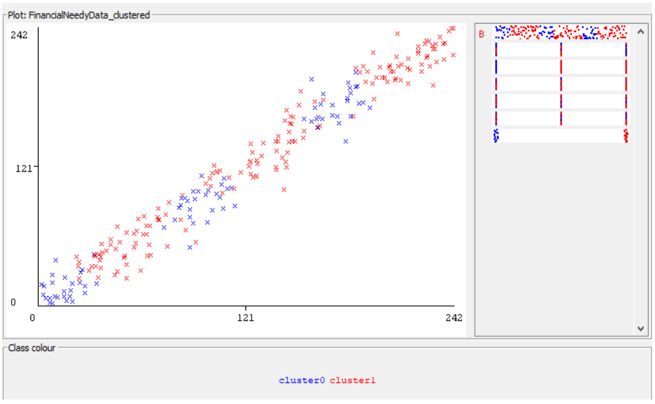

Fig. 2. EM Clustered Dataset.

From Fig. 2, the dataset coloured blue represents elements of Cluster 0 whereas the dataset coloured red represents elements of Cluster 1. The final log-likelihood value after clustering was -4.83197 which is useless since it is not compared with another log-likelihood value, though a high absolute value of log-likelihood indicates a good model.

It can be deduced from the above that the K-Means algorithm with Euclidean distance produced nine (9) more needy students than the EM algorithm. However, going through the categorization list produced by both K-Means and EM algorithms, 198 datasets were categorized the same by both algorithms leaving out 45 datasets which were categorized differently by both algorithms. Of these 198 datasets, 63 of them were categorized as Cluster 0 (needy students) and 135 were categorized as Cluster 1 (financially stable students). Applying the logical AND operator for the remaining 45 datasets categorizes them as elements of Cluster 1. Thus, there were 180 elements representing 74% of the total dataset in Cluster 1 and 63 elements representing 26% of the total dataset in Cluster 0 at the end of the study.

Could the K-Means with Manhattan distance produce better results than the Euclidean distance? To answer that question, the dataset was categorized using the K-Means algorithm with Manhattan distance as well. Given the initial centroids values for Cluster 0: 0, 0. 0, 0, 0 and Cluster 1: 2, 2, 2, 0. 0 the final clusters produced the following results for the K-Means algorithm using the Manhattan distance in Table III.

Out of the 243 datasets, 108 of them representing 44% of the total dataset were categorized as cluster 0 (needy students). On the other hand, 135 of the total dataset representing 56% were categorized as cluster 1 (financially stable students. The sum of squares within-cluster distances was 350.5.

TABLE I: CENTROID VALUES OF K-MEANS ALGORITHM AFTER CLUSTERING USING MANHATTAN DISTANCE

Attributes	Cluster 0	Cluster 1
Parents Income	0	2
Student Income	0	2
Parent Debt	1	2
Medical Cost	1	1
Siblings Education Cost	1	1

Given the sum of squares within-cluster distances as a measure of the deviation of the cluster elements, the model with the Euclidean distance is better than the model with Manhattan distance due to its lower value of the sum of squares within-cluster distances. The resultant dataset of elements in cluster 0 between K-Means with Euclidean distance and EM algorithm would have been 108 when the logical OR operator has been used. This would have been the same results obtained with the K-Means algorithm with Manhattan distance.

IV. CONCLUSION

This study uses clustering methods to determine needy students who require financial assistance. The attributes of the datasets employed in the study come from the income of both parents and the student and such recurring expenditures on non-avoidable circumstances like parent debts, medical costs, and siblings' education costs. The dataset was normalized. The clustering methods were the K-Means algorithm with Euclidean distance measure and EM algorithm. As a control measure, the K-Means algorithm with Manhattan distance was also used to categorize the dataset. The study revealed that hybrid clustering methods yield better results. Moreover, the K-Means algorithm with Euclidean distance produces better categorization than the K-Means algorithm with Manhattan distance when the data is normalized. Future work should look at using the categorized dataset as a basis for a classification model that will be free from manual review, randomness and discretion of the selection committee.

ACKNOWLEDGMENT

We wish to thank all the respondents who contributed to the success of this study by completing the questionnaire.

CONFLICT OF INTEREST

The authors declare that they do not have any conflict of interest.

REFERENCES

[1] Hu L, Zhao X, Fan S, Gui. Identification model for needy Undergraduates Based on FFM. *Applied Mathematics.* 2020; 11:8-22.

[2] Wang X, Wu S, Zhu Q. A Method of Determining Needy Students Based on Deep Learning. *Journal of Physics: Conference Series.* 2018; 1069(1): 012122.

[3] Luo X, Kong L, Wang J. Research on the Financial Difficulty Level Recognition of Needy Students Based Upon the Decision Tree C5.0. Journal of Physics: Conference Series, 2020; 1634(1): 012056.

[4] Zhu Y, Zhu L. Identification of Impoverished College Students Based on LR-RF Dual Indicators. In *10th International Conference on Educational and Information Technology (ICEIT) (pp. 195-199). IEEE*. 2021.

[5] Opoku SK, Appiah S. Automating Students' Activities in Higher Educational Institutions. *International Journal of Computer Applications Technology and Research.* 2016; 5(11): 693-697.

[6] Opoku SK. Artificial Neural Network Context-Aware Reasoning Engine Model for Predicting User-Ranked Activities. *European Journal of Electrical Engineering and Computer Science.* 2021; 5(6): 36-42.

[7] Ansong MO, Opoku SK. An Adaptive Neuro-Fuzzy Inference System for Predicting Survivability Rate in Underground Mining Accident. *International Conference on Applied Science and Technology Conference Proceedings*, 2020; 6(1): 250-265.

[8] Singh M, Kaur K, Singh B. Cluster algorithm for genetic diversity. *World Academy of Science, Engineering and Technology.* 2008; 18: 453-457.

[9] Kashyap M, Gogoi S, Prasad RK. A Comparative Study of Partion-based Clustering Methods. *International Journal of Creative Research Thoughts.* 2018; 6(2): 1457-1463.

[10] Dubey R, Kavita Y., Comparative Study of Various Clustering Methodologies. *New Horizons.* 2021; XVIII:8-19.

[11] Azhir E, Navimipour NJ, Hosseinzadeh M, Sharifi A, Darwesh A. An automatic clustering technique for query plan recommendation. *Information Sciences.* 2021; 545: 620-632.

[12] Shahidinejad A, Ghobaei-Arani M, Masdari M. Resource provisioning using workload clustering in cloud computing environment: a hybrid approach. Cluster Computing. 2021; 24(1): 319-342.

[13] Beltran B, Darnes V. Survey of Overlapping Clustering Algorithms. *Computación y Sistemas*. 2020; 24(2): 575-581.

[14] Kazi A, Kurian DT. A Survey of Data Clustering Techniques. *International Journal of Engineering Research & Technology*. 2014; 3(10): 174-17.

[15] Ruggieri A, Stranieri F, Stella F, Scutari M. Hard and Soft EM in Bayesian Network Learning from Incomplete Data. *Algorithms*. 2020; 13(12): 1-17.

[16] Opoku SK. Agent-Based Automated Negotiation System for Handling Vague Data Using Artificial Neural Network with Adaptive Back Propagation Algorithm. *International Journal of Advances in Computer Science and Technology.* 2013; 2(11): 247-255.

[17] Opoku SK, Rao DS, Ansong MO. A Context-Aware Reminder System for Tracking SMS in Mobile Device Environment Using Artificial Neural Network.*International Conference on Applied Science and Technology Conference Proceedings.* 2021; 6(1): 140-150.

[18] Sasu DD. *International Poverty Rate in Ghana from 2017-2022. Statista*. [Internet]. 2021 [cited 2021 March 17]. Available from: https://www.statista.com/statistics/1222084/international-poverty-rate-in-ghana/#:~:text=The%20international%20poverty%20rate%20in,was%20measured%20at%2011.1%20percent

[19] Kingombe C. Africa's rising middle class amid plenty and extreme poverty. European Centre for Development Policy Management Discussion Paper, 2014.

[20] Cohen A. Déclaration, La Revue juive, 2015. French.

[21] Opoku SK. An indoor tracking system based on Bluetooth technology. Arxiv [Preprint] 2012. Available from: https://arxiv.org/ftp/arxiv/papers/1209/1209.3053.pdf