JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 1, 2022, p. 111-120
https://publishoa.com
ISSN: 1309-3452

Faculty Performance and Clusteirng – A Critical Review

Preeti Jain
Research Scholar, MSIT, MATS University, Raipur, Chhattisgarh, India
preeti.jain3077@gmail.com

Dr. Gyanesh Shrivastava
Research Supervisor, MSIT, MATS University, Raipur, Chhattisgarh, India
gyaneshnkshrivastava@gmail.com

*Dr. Umesh Kumar Pandey
Co-Research Supervisor, United Institute of Technology, Naini, Prayagraj, Uttar Pradesh, India
umesh6326@gmail.com

Received 2022 March 15; **Revised** 2022 April 20; **Accepted** 2022 May 10.

ABSTRACT
Pattern analysis or grouping in the dataset has a very complex problem. Data analytics algorithms used to solve this problem are known as the clustering algorithm. The variety of data and objective of the problem for grouping the data lead to various clustering algorithms. The clustering algorithm task is to identify the identical item and place them into the respective based on distance measure or similarity index.

Over the period, Clustering got new ways to group the data. So, researcher formed a group for clustering algorithms into different categories based on specified criteria. Each category has a specific methodology, type of data, criteria to stop, and the number of clusters to be formed.

This paper studies various categories of clustering algorithms and popular algorithms into specific Clustering and critical features of the algorithm.

Keywords—Faculty Performance, supervised learning, unsupervised learning

INTRODUCTION

Humans are in quest of knowledge. Earlier, the pursuit of knowledge was satisfied by the human process. Exposure to the extensive data and processing it through human minds

Extracting patterns or trends from the data is a complex and calculation-intensive task with human processes. However, many applications like stock market research, image-based data grouping, sentiment trends need data grouping. Clustering is one of the methodologies to form the group.

Data converted itself into big data. Even humans are trying to group high-dimensional data. Nowadays, clustering algorithms incorporate various concepts and complex mathematical to group the data. The nature-based algorithm is also in discussion of the group data item. This led to a long list of clustering algorithms.

The long list of Clustering is categorized into various categories depending on the core concept clustering algorithm used to group the data item.

This research paper talks about algorithms falling into the various categories and core concepts that make any specific clustering algorithm different from others. The Paper also represents an important evaluation index to measure the performance of the clustering algorithm.

Data Mining and faculty performance evaluation

Hassan et al. [15] assessed that Qatar's universities and colleges have three primary appraisal systems: Teaching, Scholarly endeavor, and Service to the university/College. For achieving the desired excellence, all components should have harmony. They have developed the Structural Equation Model approach for finding the relationship between these three components. They see a significant difference between Teaching and Service and no academic endeavor support with either teaching or service performance [16] [17].

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 1, 2022, p. 111-120
https://publishoa.com
ISSN: 1309-3452

The OECD, 2009 [16] study report said that appraisal and feedback increase job satisfaction and a lesser degree of teacher's job security. Also, it is fair and valuable to teacher development. Moreover, the teacher reports that appraisal and feedback have contributed to their development as teachers suggest that such systems improve school improvement. According to teachers' reports of their impact, strengthening teacher appraisal and feedback can develop schools' teaching skills.

Policymakers' numerous initiatives to improve schools have had teacher development at the core [18].

Teachers' feedback on their performance helps them better shape and improve their teaching practice and support effective school leadership to develop professional learning communities. Simultaneously, teachers should be accountable for their performance and progress based on demonstrated effective teaching practice. [16]

Developing a comprehensive approach may be costly but is critical to conciliate the demands for educational quality, enhance teaching practices through professional development, and recognize teacher knowledge, skills, and competencies.

Teachers are critical to raising education standards, so teachers are considered a significant resource in the educational environment. Improvement of school efficiency and equity depends, in no small measure, on ensuring that teachers are highly skilled, well resourced, and motivated to perform at their best. Raising teaching performance is perhaps the policy direction most likely to lead to substantial learning gains [18].

School evaluation and teacher appraisal and feedback systems aim to maintain standards and improve student performance. There will likely be significant benefits from the synergies between school and teacher evaluations. Therefore, school evaluation focus should be linked to or affect teacher evaluation priority to achieve the most significant impact [16].

By contrast, teachers and their unions expect social recognition of their work and opportunities for professional growth by developing a formative teacher evaluation system [19].

In their research, Avalos and Assael [19] grouped teacher's responsibilities into four significant areas – Planning and Preparation, Classroom Environment, Instruction, and Professional Responsibilities. Each of these components consists of several elements to evaluate. The proposed framework influenced many teacher evaluation systems around the world. For instance, the framework inspired Chiles's four domains and twenty assessment criteria.

Anup K Ghosh et al. [20] proposed a model and a flexible demonstration compared to other models. It grabs the decision maker's confusion and minimizes the vagueness in human (expert) decision-making. The imprecision of human judgment nullified using linguistic variables. All the feedback took to keep the view of an optimistic, most likely, and pessimistic decision-making environment. The model presented in this work also takes care of crisp inputs for evaluating faculty performance and fuzzy criteria weights to make the overall performance evaluation more realistic. This method applies to measure faculty performance in the different fields of education with greater efficiency. Further, the software makes the data collection more manageable and comfortable, but the outputs are error-free and easy to interpret.

Akbar Jesarati et al. [21] used a descriptive survey and multistage sampling method for data collection and applied it to four randomly selected faculty of the Islamic Azad University of Tabriz. They identified the order of faculty performance evaluation factors.

Gorji and Siyami [22] opined that faculty and evaluation criteria' performance has a healthy relationship.

Rajabi and Popzan's [23] research suggests integrating qualitative and quantitative design tools for faculty performance evaluation.

Malekashahi et al. [24] study identified that most responses favored workshops.

Sivasankari S et al. [25] proposed semantic web architecture for online feedback systems and generating reports using the automatic process. The advantage of the system highlighted was the removal of manual effort easy management.

Sampson J P et al. [26] clarified the elements responsible for faculty performance success. They also opined that faculty vision must be clear for successful faculty performance.

Las Johansen B Caluza et al. [27] recommended giving attention to professionalism, subject knowledge, commitment, and teaching for independent learning.

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 1, 2022, p. 111-120
https://publishoa.com
ISSN: 1309-3452

Jyoti G et al. [28] work use a fuzzy expert system to assess faculty performance. They opined that an expert system is suitable for quantitative and qualitative facts about the faculty. The fuzzy expert system model converts the qualitative value into a numeric value. Qualitative and quantitative data evaluation encourages faculty members' satisfaction, quality, and efficiency.

Mustafa A. [29] discussed instructor performance prediction. Finally, Priyanka R Shah et al. [52] discussed distributed data mining to predict faculty.

Moghtodia Leila et al. [30] discussed performance study according to higher education faculty's talent management approach.

R K Banu et al. [31] discussed using the NLP approach to assess the faculty's success pattern.

Rand kh Hemaid et al. [32] and Thy Van et al. [33] used a data mining approach to improve the faculty.

Ajay Kumar Pal et al. [34] research work talk’s regular faculty assessment. The studied scale based value on various teacher performance evaluation parameters. Their research studies four different data mining methods, i.e., naïve Bayes, ID3, classification and regression tree (CART), and logical analysis of data (LAD). Out of these four naïve Bayes classification methods have been identified as the best method for the dataset and generates the slowest average error compared to others.

Archana Bhardwaj et al. [35] research compares single classifier and multi classifier approaches on faculty performance. The finding of the research work highlights the importance of the multi classifier compared to the single classifier.

Asanbe M O et al. [36] research work studies six attributes to find the best attribute to identify the faculty performance. Three algorithms, i.e., ID3, C4.5, and MLP used to analyze the data and find that work experience has the highest important factor in performing the faculty.

Nirmala G et al. [37] research work studied k means clustering algorithm to form faculty members group on evaluation parameter considered for the research purpose and apriori association rule to find the most qualified faculty for activities based on minimum support level.

Priti Ughade et al. [38] research work considers student complaints about faculty, student review feedback, student feedback, and student results. The data was studied using K nearest neighbor.

Renuka Agrawal et al. [39] research work proposes an optimal algorithm and framework of faculty evaluation.

SUPERVISED AND UNSUPERVISED LEARNING

Grouping data into the absolute number of groups is possible in two ways, i.e., unsupervised (Clustering) or supervised (classification) learning. Supervised learning requires a well-defined class label and range of values to map into the respective class. The limitation of supervised learning is that levels of each group are defined before being mapped. Defining each group's levels needs well-proven research or a researcher's biased level because of researchers' perception and belief. Secondly, some stories are never mapped with any data because no data qualify for mapping to that class. On the other side, unsupervised learning groups data into a pre-decided number of groups. Since no pre-defined level is needed, unsupervised learning is preferred in the scenario where the level of any group is not available or the scenario in which the researcher groups data into a specific group. Unsupervised learning is beneficial because each group has a few numbers.

Effectiveness of the supervised learning algorithm tested using confusion matrix and particular statistical functions, i.e., precision, recall F score. In addition, unsupervised learning effectiveness is measured using various cluster validation indexes, i.e., partition entropy and silhouette index.

CLUSTERING TECHNIQUES:

Clustering techniques group data on specific criteria referred to as measures. All measurement methods are grouped into two categories, i.e., Euclidean space and non-Euclidean space. Euclidean space uses numerical values to find the distance between the data points to form the group. Non-Euclidean space uses features of the data to find the distance. These distances belong to Euclidean or Non-Euclidean spaces, representing the similarity or dissimilarity between the data points. Similar data points are placed into the same group, whereas dissimilar data points are placed into different groups. For example, the euclidean distance measure finds the shortest distance, and the Manhattan distance measure is useful when path-based distance needs to be calculated.

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 1, 2022, p. 111-120
https://publishoa.com
ISSN: 1309-3452

Some researchers also talk about Mahalanobis, Minkowski, and Pearson correlation to find similarities in Euclidean space.

Non-Euclidean space distance measure techniques include Jaccard similarity and Hamming similarity. Jaccard similarity uses the concept of intersection and union size for finding the similarity score. Hamming distance finds similarity based on binary series of values, and hamming distance value is the difference of bit at the same position. Many more non-Euclidean spaces similarities are proposed for the distance measure.

One of the comprehensive works on the survey of clustering technique was conducted by Dongkuan Xu and Yingjie Tian[1].

Over the period, unsupervised learning various concepts to form the group.

Partition-based clustering method initially asks the number of clusters in which the entire data set needs to be partitioned. Each cluster has its centroid value and data items placed into the nearest centroid cluster. T. Velmurugan and T. Santhanam [2] carried out a comprehensive study on partition-based Clustering. In all partition-based clustering algorithms number of clusters needs to be provided at the initial stage. While working with the initial data value of K affects the result and runtime of the algorithm.

K means Clustering is a popular method for partition-based Clustering. This algorithm works for the Euclidean space problem. K means Clustering is sensitive for the outliers and easy to implement. The centroid value, symbolic value of the specific group, find by mean central tendency.

K medoid is approach works similar to K means, but the difference is that centroid value is medoid value rather than mean value. Medoid value is one of the data elements of that specific group. Medoid value represents the central location and is calculated using Manhattan distance measure rather than Euclidean distance measure. K medoid is not sensitive for the outlier.

Portioning around medoid (PAM) algorithm works on the concept of medoid. This algorithm works in two phases. Phase one is known as BUILD, and the second phase is known as the SWAP phase. The BUILD phase partitions the data set into the K group, and the SWAP phase exchanges the element to improve the cluster quality. This algorithm is also not sensitive to outliers. Therefore, PAM is effective for the small data set. However, the PAM algorithm is poor for scalable implementation.

Kaufman and Rousseuw proposed Clustering LARge Application (CLARA) as an extension of K medoid and used the concept of PAM. In addition, CLARA supports scalability, which was the limitation of the PAM algorithm.

CLARANS (Clustering Large Application based upon RANdomized Search) algorithm uses graph searching concept and assigns cost to each node of dissimilarity. CLARANS is much better than PAM and CLARA on the front of scalability and efficiency.

Hierarchical Clustering has an entirely different concept than partition-based Clustering. Instead, it uses the tree's concept to represent the groups of data elements. The popular visualizing technique for hierarchical technique is dendrogram.

The hierarchical clustering core concept is tree-based grouping. Data elements are grouped level by level based on similarity. Hierarchical clustering algorithms follow two basic approaches for forming the tree, i.e., top-down and bottom-up. The top-down approach-based hierarchical technique is divisive (DIANA), whereas the bottom-up approach-based Clustering is agglomerative (AGNES). MONolethic Analysis (MONA) is a divisive hierarchical clustering algorithm handled with binary data. The resultant cluster represents the individual element. The agglomerative clustering technique starts with a data element and ends with one cluster representing all elements into that cluster.

Agglomerative Clustering is based on links, i.e., single link (SLINK algorithm) and complete link (CLINK). SLINK works on the concept of nearest neighbor, whereas complete link uses the concept of farthest link for clustering purposes.

Balanced Iterative Reducing and Clustering using Hierarchies (BIRCH) introduced by Tian Zhang, Raghu Ramkrishnan, and Miron Livny[4]. BIRCH starts with distribution information. Then, this distribution information is used to build the hierarchical tree. This handles outlier effectively and supports linear scalability.

Sudipto Guha proposed Clustering Using REpresentativee (CURE) [5]. The core concept of CURE is representative points of cluster. It starts with random sampling and equal partition size and then removes outliers

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 1, 2022, p. 111-120
https://publishoa.com
ISSN: 1309-3452

to form the cluster. The limitation of the CURE algorithm is a failure in handling different densities and noise of data. Nevertheless, the CURE algorithm can handle large datasets.

RObust Clustering using linKs (ROCK) [6] hierarchical clustering algorithm specifically developed for the categorical and Boolean dataset. ROCK includes nearest neighbor and relocation concept for creating an agglomerative approach to build the tree. Moreover, ROCK faces the problem with incremental data set.

George Karypis, Eui-Hong Han, and Vipin Kumar proposed Chameleon[7]. Chameleon works in two phases. In the first phase, the graph partitioning algorithm is used with the k nearest neighbor, and in the second phase, portions are merged to form a sub-cluster. Unfortunately, chameleon performance is inferior for high-dimensional datasets [3].

When data points are placed into a specific group based on likelihood or probability, Fuzzy clustering algorithms are preferred. A fuzzy clustering algorithm assumes that any item can belong to any group. Only probability defines that actually from which group data point belongs. A fuzzy clustering algorithm works very well where data points overlap each other. Fuzzy C means (FCM) is a soft fuzzy clustering method. Fuzzy clustering algorithm uses the concept of fuzzy approach to group data points. Fuzzy C Means (FCM) is one of the most popular algorithms of this category and similar to K means partition algorithm, except data points have membership grade to represent the probability of mapping into the specific cluster. The popular fuzzy concept-based algorithm, i.e., Fuzzy Compactness and Separation (FCS) and Mountain Method (MM).

Distributed clustering algorithms are well suited for spatial points. These spatial points have arbitrary shapes and sizes. Spatial database problems need special clustering techniques because of the feature of the space, viz. earth surface. Distribution Based Clustering of LArge Spatial Database (DBCLASD) is one of the popular algorithms in this category. Distributed Based Clustering of LArge Spatial Databases (DBCLASD) has been introduced to address arbitrary distributed data point’s problem. DBCLASD works well with uniformly distributed data and is well suited for high dimensional feature space [8]. Gaussian Mixture Model (GMM) is another algorithm of this category. Gaussian Mixture Model (GMM) is another distributed clustering algorithm. GMM clustering algorithm uses the concept of expectation-maximization and probability of each point to form the cluster.

Density-based clustering algorithms make clusters based on size and density. Density-Based Spatial Clustering of Applications and Noise (DBSCAN) is one of the popular clustering algorithms of this category. DBSCAN takes inputs from the user, i.e., specify the number of points and the region's radius. Noise points in the DBSCAN algorithm are eliminated from the dataset. Ordering Points To Identify the Clustering Structure (OPTICS) density-based algorithm works with an arbitrary selection of points, and reachable points are grouped into one cluster; otherwise, respective clusters are created or assigned for the non-reachable cluster. The mean-shift algorithm is another density-based algorithm that uses the concept of the centroid. This centroid shifts when a new data point adds to the specific cluster. One crucial point is that prior knowledge of clusters and shapes is not required.

Graph clustering algorithms use the concept of vertex and edges to form the cluster. Graph clustering identifies groups or clusters considering the structure of the edges. The graph clustering algorithm is widely applied in data transformation, information networks, database systems, biological and sociological networks, and many more similar problems where data points are connected and can be represented in the vertex and edges representation. CLICK, and MST is two popular algorithms in the graph clustering algorithm.

Some data points are in a multi-resolution data structure which forms a grid. The clustering algorithm used for Clustering these points is considered a Grid clustering algorithm. Unlike other clustering algorithms, the grid algorithm's complexity depends on the grid's structure rather than data points. STatistical INformation Grid (STING) algorithm is one of the popular algorithms in grid-based Clustering. STING algorithm addresses spatial database problems for two-dimensional data only [9]. CLIQUE is another algorithm that works with the grid approach. CLIQUE divides the dataset into a subset and discriticizes the data to form a grid and cluster identified by counting the data points falling in one grid.

Model-based clustering algorithms are very different in comparison to other clustering algorithms. Model-based algorithm prefers to use a mathematical model to fit the data and group them into the cluster. COBWEB clustering is one of the most popular algorithms in this category. COBWEB uses incremental conceptual

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 1, 2022, p. 111-120
https://publishoa.com
ISSN: 1309-3452

learning concepts in unsupervised learning groups. COBWEB forms a hierarchical clustering tree as observed in the classification problem. The limitation of the COBWEB is that it is unable to work with a large dataset. Self-Organizing Map (SOM) is another model-based clustering algorithm. SOM uses an artificial neural network concept for unsupervised learning. SOM uses topological properties of data points and applies competitive neural network learning. SOM is unable to handle mixed data, i.e., categorical, discrete. Adaptive Resonance Theory (ART) is the next model-based algorithm that uses neural network concepts to build the model. The ART clustering algorithm is self-organizing and uses a competitive learning approach to build the cluster. Add-on in the ART algorithm is to find new patterns and include the experience of the data in finding new patterns. ART has different variations. The ART variation, i.e., ART1, ART2, ART3, are used for unsupervised learning. Since ART uses experience to find new patterns, the algorithm's consistency widely depends on the order in which training data is provided to the algorithm.

The kernel-based clustering algorithm is different from the traditional algorithm. Kernel-based Clustering adopts an approach so that appropriate cluster forms from the dataset. Ensemble clustering is the modern method in this category. However, it has been observed that any clustering algorithm is inappropriate for all kinds of data because the dataset has different distributions, types, sizes, etc. Ensemble clustering algorithm uses multiple clustering algorithms to solve the problem. Ensemble clustering algorithms objective is to provide a better result with consistency and quality. Ensemble clustering algorithms use similarity measure, similarity measure to for its performance.

Swarm intelligence is also a kernel-based algorithm and bio-inspired algorithm. Swarm intelligence clustering algorithm developed for large data set Clustering with the help of a fewer number of prototype cluster. Swarm intelligence clustering is very popular for real-world clustering problems [10].

Quantum theory-based kernel clustering algorithm is helpful in renewable energy, complex physical simulation, biological research, etc.

Spectral graph clustering is also known as self-tuned Spectral Clustering. Spectral Clustering determines the number of clusters and creates an affinity matrix using a similarity matrix.

In contrast to traditional Clustering, Affinity propagation does not require the number of clusters. Affinity propagation's core concept is to find relative attractiveness for the sender to target points used to prepare the affinity matrix. Affinity matrix utilized to form the cluster.

In the age of big data, the data stream is pervasive. Various data generation sources generate a speedy and massive amount of relevant data. Datastream clustering algorithm evolved to address this kind of data. Datastream clustering algorithm needs to be robust while dealing with the outlier and noisy data set.

EVALUATION PARAMETERS OF CLUSTERING ALGORITHM:

Clustering algorithms' performance was tested over many parameters. Figure 3.1 represents various clustering performance measures. Evaluation indices are classified into two categories internal and external evaluation metrics. Internal evaluation metrics were used for testing the validity of the algorithm. External evaluation metrics are used to test any algorithm with external data.

The number of internal and external evaluation indices proposed, out of which three internal evaluation metrics named in figure 1 are Davis Bouldin, Dunns Indicator, Silhouette. In contrast, four external evaluation metrics are Rand, Jaccard, Foulkes-Mallows, Mutual Information, and confusion matrix.

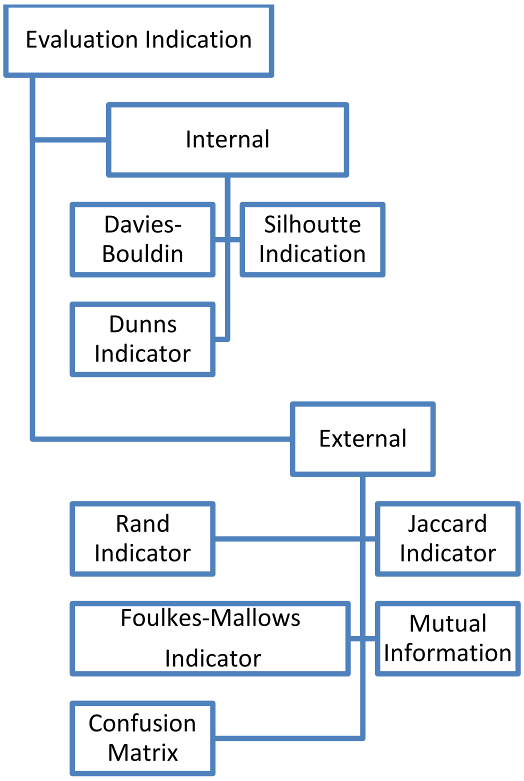

Figure 1 Evaluation Indicators for Clustering Algorithm

Davis boulding score measures the average similarity of each cluster. This similarity is the ratio of within-cluster and between cluster distances. Lower values of this score indicate better Clustering [11].

Dunns Index is another internal metric of the cluster. Higher Dunn's vale indicates better Clustering. Dunn's index score depends on the compactness and less variance cluster. Dunn's index is the ratio of min separation and maximum diameter.

The silhouette index is the next internal metric to test the performance of the clustering algorithm. The objective of the silhouette index is to find the distance and express data closeness in the cluster compared to the neighboring cluster. The range of silhouette scores is 1 to -1, representing a good to a bad cluster. Silhouette index measures the proportion of inter and intracluster distances.

William M Rand [12] proposed a rand indicator or rand indexed to measure the quality of a cluster. Rand index is used to find the accuracy when the dependent variable is not present in the data set. Rand value ranges between 0 to 1, in which 0 represents the worst pair, whereas 1 represents the best pair and value lying between 0 to 1 represents the intensity of pair goodness. Unfortunately, the rand index is not a proper unsupervised method; thus, it is less valuable.

Paul Jaccard [13] proposed Jaccard Indicator as a measure of similarity coefficient in the data. Jaccard coefficient is popularly used to measure the cluster goodness for the data which belongs to binary or categorical data.

Foulkes-Mallows [14] indicator finds similarity between two clusters and engages confusion matrix for finding the similarity value. Foulkes–Mallow's indicator uses a positive predictive rate and a true positive rate. Foulks-Mallows indictor is popularly used with hierarchical clustering goodness measure.

Mutual information uses the concept of entropy that measures randomness in the data. Normalized mutual information shows the reduction in the entropy of cluster numbers. Sometimes normalized information is considered as information gain, but it is incorrect. Information gain adds the impact of one level to another level, but normalized information gain has no such implications in measuring the quality of the cluster.

CONCLUSION

The discussion present in the research paper elaborates on key features of the popular clustering algorithm. The entire discussion shows that the clustering algorithm requires either number of clusters or the size of radius or density into specific Clustering from the user. Specifying the value of either number of clusters or size of radius or density into specific Clustering is user-dependent. Thus new methods and algorithms are needed which find a parameter to define the number of the cluster with the help of data.

Funding: Authors did not receive any funding for this research work.

Conflict of Interest: All Authors declares that none of the author has any conflict of interest.

Ethical approval: This article does not contain any studies with human participants or animals performed by any of the authors.

References:

1. Dongkuan Xu and Yingjie Tian, “A Comprehensive Survey of Clustering Algorithm”, Ann. Data. Sci. (2015) 2(2):165–193 DOI 10.1007/s40745-015-0040-1
2. T. Velmurugan and T. Santhanam, 2011. A Survey of Partition based Clustering Algorithms in Data Mining: An Experimental Approach. Information Technology Journal, 10: 478-484. **DOI:** 10.3923/itj.2011.478.484
3. Dabhi Deepank P and Patel Mihir R, “Extensive Survey on Hierarchical Clustering methods in Data Mining”, International Research Journal of Engineering and Technology, Vol 3 Issue 11 Nov 2016.
4. Tian Zhang, Raghu Ramakrishnan, Miron Livny, “BIRCH: A New Data Clustering Algorithm and Its Applications”, ACM, Springer, Year: 1997.
5. Sudipto Guha, Rajeev Rastogi, Kyuseok Shim,” CURE: an efficient clustering algorithm for large databases”, ACM, Year: 1998.
6. Sudipto Guha, Rajeev Rastogi, Kyuseok Shim,” ROCK: A Robust Clustering Algorithm for Categorical Attributes”, Year: 1999.
7. George Karypis, Eui-Hong Han,Vipin Kumar,” CHAMELEON: A Hierarchical Clustering Algorithm Using Dynamic Modeling”, IEEE,Year: 1999.
8. Xiaowei Xu, Martin Ester, Hans-Peter Kriegel, Jörg Sander, “A Distribution-Based Clustering Algorithm for Mining in Large Spatial Database”, Published in the Proceedings of 14th International Conference on Data Engineering (ICDE’98)
9. Guojun Gan, Chaoqun Ma and JIanhong Wu,” Data Clustering: Theory, Algorithms, and Applications”, Soc Ind Appl Math (2007) https://doi.org/10.1137/1.9780898718348.ch12
10. Abraham A., Das S., Roy S. (2008) Swarm Intelligence Algorithms for Data Clustering. In: Maimon O., Rokach L. (eds) Soft Computing for Knowledge Discovery and Data Mining. Springer, Boston, MA. https://doi.org/10.1007/978-0-387-69935-6_12
11. Davies, David L.; Bouldin, Donald W. (1979). “A Cluster Separation Measure”. IEEE Transactions on Pattern Analysis and Machine Intelligence. PAMI-1 (2): 224-227
12. W. M. Rand (1971). "Objective criteria for the evaluation of clustering methods". Journal of the American Statistical Association. American Statistical Association. 66 (336): 846-850. Doi:10.2307/2284239. JSTOR 2284239.
13. Jaccard, Paul (February 1912). "THE DISTRIBUTION OF THE FLORA IN THE ALPINE ZONE.1". New Phytologist. 11 (2): 37–50. doi:10.1111/j.1469-8137.1912.tb05611.x. ISSN 0028-646X.
14. Fowlkes, Edward B; Mallows, Colin L (1983). "A method for comparing two hierarchical clusterings". Journal of the American Statistical Association. 78 (383): 553–569. doi:10.1080/01621459.1983.10478008.
15. Hassna L O, Raza S, "An assessment of the relationship between the faculty performance in teaching," scholarly endeResearch in Higher Education Journalavor, and service at Qatar University, pp. 1-18.

16. OECD 2009, Creating Effective Teaching and Learning Environments: First Results from TALIS, School Evaluation, Teacher Appraisal, and Feedback and the Impact on Schools and Teachers,– ISBN 978-92-64-05605-3, pp 138-139
17. OECD Review on Evaluation and Assessment Frameworks for Improving School Outcomes, Teacher Evaluation A Conceptual Framework and examples of Country Practices, December 2009, pp 1-37
18. OECD (2005), Teachers Matter: Attracting, Developing, and Retaining Effective Teachers, OECD, Paris.
19. Avalos, B. and J. Assael (2006), "Moving from resistance to agreement: The case of the Chilean teacher performance evaluation," International Journal of Educational Research, Vol. 45, No. 4-5, pp. 254-266.
20. Anup K Ghosh, Debmallya Chatterjee, Biswarup Ghosh, A conceptual framework of faculty performance evaluation, Asian Journal of Management Research, ISSN 2229-3795, pp. 217-229.
21. Akbar Jesarati, Nasser Rahmat, Salar Ghanbari Erdi, Hossein Babazadeh, Yosef Hajnazary and Asgar Jesarati, An investigation of performance evaluation index from the perspective of Islamic Azad University of Tabriz faculty members in 2012, Euro. J. Exp. Bio., 2013, 3(4):165-172
22. Gorji, M and S. Siyami, Journal of Knowledge Management. 2007, 5 , 11
23. Rajabi, S. and A. Pop Zan, Journal of Higher Education, 2010, 3, 1.
24. Malekshahi, F. Sheikh. Ali and MJ Tarrahi, J of Nursing Research. 2010, 5,19.
25. Sivasankari.S, Srimathi. P.S, Ramya.S, Dr.G.Fathima, Online Feedback System for Educational Institutions for Better Evaluation of Faculty's Performance Using Semantic Web (S.W.) Technology International Journal of Innovative Research in Science, Engineering and Technology An ISO 3297: 2007 Certified Organization Volume 5, Special Issue 2, March 2016 ISSN (Online): 2319 - 8753 ISSN (Print): 2347 – 6710
26. James P. Sampson, Jr., Marcy P. Driscoll, David F. Foulk, and Pamela S. Carroll, Successful Faculty Performance in Teaching, Research and Original Creative Work, and Service, Florida State University 2010 pp 1-17
27. Las Johansen B. Caluza, Devine Grace D. Function, Rommel L. Verecio Jeffrey C. Cinco, Lowell A. Quisumbing, Mark Lester Laurente, Micheline A. Gotardo, Faculty Performance Evaluation in A Philippine University Information Technology Program, IOSR Journal Of Humanities And Social Science (IOSR-JHSS) Volume 22, Issue 9, Ver. 8 (September. 2017) P.P. 28-36 e-ISSN: 2279-0837, p-ISSN: 2279-0845. www.iosrjournals.org DOI: 10.9790/0837-2209082836 www.iosrjournals.org pp 28-36
28. Ms. G. Jyothi, Mrs. Ch. Parvathi, Mr. P. Srinivas, and Mr. Sk. Althaf Rahaman Fuzzy Expert Model for Evaluation of Faculty Performance in Technical Educational Institutions, Journal of Engineering Research and Applications www.ijera.com ISSN: 2248-9622, Vol. 4, Issue 5(Version 7), May 2014, pp.41-50
29. Mustafa Agaoglu, "Predicting Instructor Performance Using Data Mining Techniques in Higher Education," Digital Object Identifier 10.1109/IEEE ACCESS.2016.2568756.
30. Moghtadia Leila and Taji M., "Study of the Performance of Faculty Members According to Talent Management Approach in Higher Education," Academic Journals Vol 11(8) pp 781-790, 23rd April 2016 ISSN 1990-38
31. R K Banu and R Ravanan," A Competency Framework Model to Assess Success Pattern for Indian Faculties A NLP Based Data Mining Approach," Journal of Scientific and Industrial Research, Vol 75,November 2016, pp. 662-666.
32. Randa Kh. Hemaid, Alaa M. El-Halees, "Improving Teacher Performance using Data Mining," International Journal of Advanced Research in Computer and Communication Engineering Vol. 4, Issue 2, February 2015.
33. Thuy-Van T.Duong, Thuc-Doan Do, Ngoc-Phien Nguyen," Exploiting Faculty Evaluation Forms to Improve Teaching Quality: An Analytical Review," Science and Information Conference 2015, July 28-30, 2015 | London, UK.
34. Ajay Kumar Pal, Saurabh Pal (2013), "Evaluation of Teacher's Performance: A Data Mining Approach," A Monthly Journal of Computer Science and Information Technology ISSN 2320088X, IJCSMC, Vol. 2, Issue. 12th December 2013, pg.359 – 369.
35. Archana Bhardwaj, Mamta Bhusry, "Two Approach Comparisons for Relative Evaluation of Faculty Performance Using Data Mining Techniques," IEEE, 978-1-4799-3064- 7/14/$31. 00©20 14.
36. Asanbe M.O., Osofisan A.O., William W.F., "Teachers' Performance Evaluation in Higher Educational Institution using Data Mining Technique," International Journal of Applied Information Systems (IJAIS) –

ISSN: 2249-0868 Foundation of Computer Science FCS, New York, USA Volume 10 – No.7, March 2016 – www.ijais.org.

37. Nirmala G, P.B. Mallikarjuna," Faculty Performance Evaluation Using Data Mining "International Journal of Advanced Research in Computer Science & Technology (IJARCST 2014), ISSN: 2347 - 8446 (Online) ISSN: 2347 - 9817 (Print) Vol. 2, Issue 3 (July - Sept. 2014).
38. Priti Ughade, S.W.Mohod," A Survey on "Analysis of Faculty Performance Using Data and Opinion Mining," International Journal of Innovative Research in Computer and Communication Engineering, ISO 3297: 2007, Vol. 4, Issue 1, January 2016.
39. Renuka Agrawal, Jyoti Singh, A.S. Zadgoankar," Formative Assessment for Performance Evaluation of Faculty using Data Mining", proceedings of IRF international conference, 31st July, 2016, Ernakulam, india, isbn: 978-93-86083-78-4.

AUTHORS PROFILE

Ms Preeti Jain is research scholar at MATS University Raipur Chhattisgarh. Ms Jain achieved MS in Computer Science from M C R P V Bhopal and MCA from Dr C V Raman University Bilaspur Chhattisgarh. Ms. Jain is actively engage in research work. The research interest area of Ms Jain is data mining.

Dr. Gyanesh Shrivastava is having rich experience of research work. Presently he is supervising eight research scholars pursuing Ph.D. in Computer Science and Applications. He has authored many book chapters and research reference books. Gyanesh has got patent publications too. He has authored more than 30 research articles in high reputed journals including many SCI publications and SCOPUS indexed publications. Gyanesh is reviewer of many journals of high impact factors. He is editorial board member of four journals. He has evaluated many Ph.D. thesis of more than five universities and conducted final viva-voce to award doctorate in philosophy. Gyanesh has delivered key note speeches/expert talk in many international conferences.

Dr Umesh Kumar Pandey works as Professor in Computer Science Department at United Institute of Technology Naini Prayagraj. 4 PhD research scholar successfully awarded PhD under the supervision of Dr Pandey. Dr Pandey has 1 patent granted and 3 published. Dr Pandey authored more than 20 research paper in Scopus and International reputed journal. Dr. Pandey has 13 years of teaching experience and 8 year of research supervision. Dr, Pandey has research exposer in the field of data mining and engage in innovative models to solve the real world problem.