



# INTERNATIONAL RESEARCH JOURNAL ON ADVANCED SCIENCE HUB

e-ISSN : 2582 - 4376  
Open Access

## RSP SCIENCE HUB

(The Hub of Research Ideas)

Available online at www.rspsciencehub.com

Special Issue of First International Conference on Advancements in Engineering & Technology (ICAET- 2020)

### K-Means Clustering on the Performance Evaluation of Faculty using Data Mining Techniques

Ms Preeti Jain<sup>1</sup> Dr. Umesh kumar Pandey<sup>2</sup>

<sup>1</sup>Research Scholar, Department Computer Science, MATS University Raipur (C.G.)

preeti.jain3077@gmail.com

<sup>2</sup>Associate Professor MATS University, Raipur (C.G.)

preeti.jain3077@gmail.com<sup>1</sup>

#### Abstract

The motivation behind this paper is to give an outline of broadly utilized measurements, to examine the qualities, advantages and disadvantages of different measurements, to portray current instructive information mining rehearses, and to give rules to assessing execution models of staff. has been discovered to be reliant on various boundaries extensively going from the person's capabilities, experience, level of commitment, research exercises attempted to institutional help, monetary achievability, top administration's help and so on The models that are basic for assessing the yield of workforce range across various verticals, however the paper locations and covers the introduction of staff dependent on contribution from understudies. The other personnel introduction assessor is the regulatory element that might be a private body or an administration unit, the affiliation or the college's self-and friend resources. The boundaries fill in as standard markers for an individual and a gathering and may influence the end later on. The standard proposed in this paper is to utilize Data Mining strategies to lead pulling out and investigation of workforce results. The fundamental idea driving the utilization of Data Drawing is to group the yield of workforce on various measures subject to novel requirements and furthermore to separate the conditions between the boundaries that will assist with finding important relations between them. Basically, these binds help to arrange new dynamic patterns. The paper limits contribution from the Department of Computer Applications through qualified foundations to understudies. The examination depends on numerous highlights, and as opposed to following the ordinary methodology, the paper legitimizes the utilization of mining approaches. K-implies is a sort of non-various leveled (gathered) information grouping that endeavors', contingent upon the methods (mm) that have been pre-masterminded, to segment information into at least two classes. The k-implies technique is utilized in numerous investigations since it is quick and fit for consolidating a lot of information with an exceptionally short computation. The k-implies calculation is the easiest and most often utilized bunching strategy. This is on the grounds that K-implies can possibly aggregate huge volumes of information with sensibly brisk and powerful preparing time.

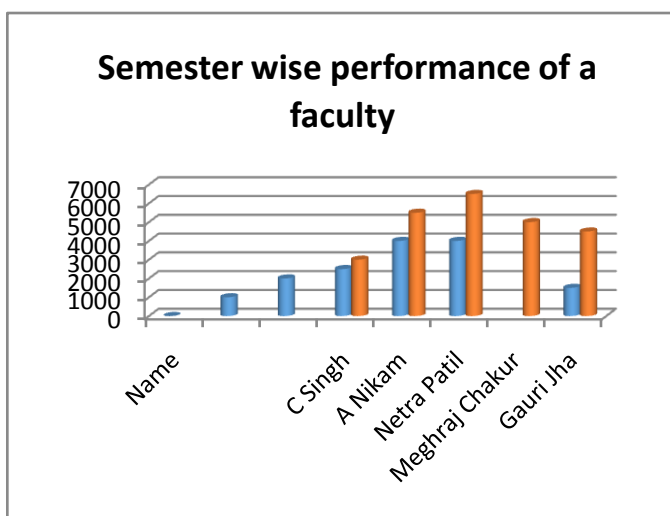
**Keywords:** Educational data mining, information Preprocessing, Analysis, Mining, Clustering, tendency Extraction, K-Means, presentation forecast, Dependencies

#### 1. Introduction

This the approach adjusted to plan the Faculty Performance evaluation and screen Structure has

been managed widely in the past papers and material boundaries perceived dependent on which the appraisal will be affirmed out. The need of Data examination and Mining in advanced

education is to mine workforce and understudies Information from an assortment of partners' point of view. In this paper we have demonstrated information pre-processing and grouping dependent on characteristic class and its sort and the reliance between the boundaries under the classification of understudy input. The staff exhibitions have likewise been separate utilizing bunching methods with respect to explicit timetables. Determine the markers, which help in amending the strategies of the organization and the scholarly height of the resources. The following is the chart portraying the equivalent and the outcomes show ordinary patterns for example with time the exhibition of an individual takes off high. The chart portrays the introduction dependent on understudy criticism for personnel and includes input for conceivable just as mandatory subjects. The chart below is a simple task of a couple of aptitudes, and a run of the mill development is likewise found in different previews. To extricate wise data and to perform compound figuring's, pattern examination and convoluted information displaying, and revealing, support was required for the usage of mathematical investigation and mining calculations. The need was to perceive significant insights regarding the less reasonable information that would assist the institute with settling on better decisions. The utilization of calculations during the time spent grouping for the most part relies upon the current information and the ends to be found. The K-Means technique is frequently utilized for this reason, making the accompanying principles[1-5]



**Chart 1. Semester wise performance of a faculty**

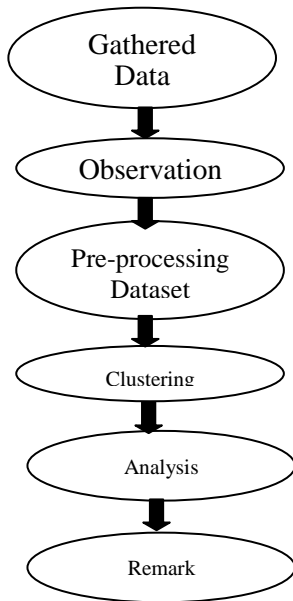
## 2. Recommender Systems

Data about substance is likewise valuable for the executives of the substance of informative frameworks – planning of new articles, channel of improper things, arrangement of clarifications, and clue messages. Data about thing likeness and bunches can be likewise material for educators as it can give them an inspiration to "live" thought in class. The setting of recommender frameworks is in numerous perspectives fundamentally the same as learning frameworks – in the two cases we have clients and things, only rather than "introduction" (the rightness of answers, the speed of answers) recommender frameworks consider "appraisals" (how much a buyer prefers a thing).

There are two crucial ways to deal with managing thing similitude's and getting segments: a "model based advancement" and a "thing closeness progress". The essential thought of the model based methodology is to develop a rearranged model that clarifies the noticed information. In light of a layout of students' responses to things we manufacture a model that foresees these answers. Commonly, the proliferation doles out various clandestine aptitudes to students and utilizations a planning of things to comparing inert elements. This sort of models can regularly be clearly spoken utilizing medium duplication, i.e., fitting a model prompts environment factorization. When we fit the model to information, things that have a similar estimation of an idle factor can be indicated as "comparable". This methodology drives normally to numerous information parts per expertise. The model is commonly registered utilizing some improvement strategy that drives just to neighbourhood optima (e.g., slope plunge). It is consequently important to address the part of introduction, and boundary setting of the inquiry system. This technique is utilized for the execution of synergistic separating in suggestion frameworks.

## 3. Research methods

The steps that must be done starting from the problem found to get a solution of the problem. Following is the framework of the stages of research used in conducting this research.



**Fig.1 Research Outline Framework**

**A. Observation**

At this stage, data collection was carried out on each lecturer who was actively teaching, who would be grouped through filling out questionnaires by students. After getting the data, the results of the data are inputted to the application.

**B. Pre-processing Dataset**

At this stage the dataset to be processed is 10 indicator indicators which are divided into 3 aspects, to determine the performance of lecturers. The 10 indicators are indicators selected from the Academic Administration Handbook for Lecturers because they are in accordance with the material that will be used as a case. Before indicator data is processed into clusters, it is necessary to do data cleaning and data transformation first. Indicator data is changed in numerical or numeric form, so that we can get the means (means) on each indicator.

**Table 1.Example of datasets that are ready to be processed into clusters**

Name of professor Initials	grounding	learn Process	capability
IIF	10.56	17.61	7.04
UUD	10.53	17.55	7.02

SSA	10.62	17.7	7.08
FF	10.69	17.81	7.12
IIW	10.48	17.47	6.99

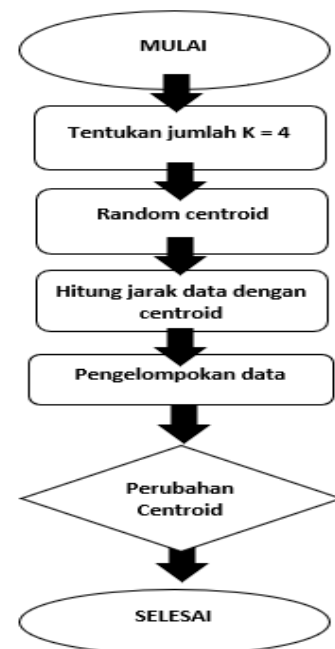
**C. Clustering**

As explained in the introduction above, clustering has used the k-means approach a great deal. K-means grouping data on the basis of the nearest distance between the evidence and the centroid measure. In measuring the distance between data and centroids, there are many methods that can be used, including the Euclidean Distance, the Manhattan Distance, and the Minkowski Distance. In this research, the author will use the Euclidean Distance calculation because it provides the best computation results [5-8]. The following is the formula for calculating

Euclidean Distance:

$$Dist(xy)=\sqrt{\sum (X_{ik} - X_{jk})^2} \quad m, k=1, \dots, n \quad (1)$$

Note: Dist (xy) = Distance between x and y  
 $X_{ik} - X_{jk}$  = variation between data and centroid  
 Clustering using the k-means method is generally illustrated in Figur. with the basic algorithm as have to be defined. Do not use abbreviations in the title or heads unless they are unavoidable.

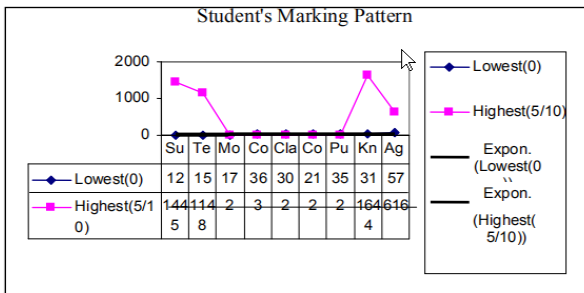


**Fig.2. Clustering**

**D. Data Clustering and Analysis**

The understudy criticism information was exposed to a channel where all the numeric traits were

changed over to the ostensible structure and the bunching calculation was applied to the person just as the characteristic/boundary set and different group insights were found. Some intriguing outcomes were acquired from the investigation of



the various bunches, as appeared in the table beneath. The quantity of students against the most minimal and most elevated worth set for every boundary considered is appeared in fig. The allowance was that understudies had given the subject information and information due weight age past the prospectus (the two boundaries) and didn't believe different boundaries to be of extraordinary criticalness and significance.

<b>Classes to Clusters</b>	
448	very high-quality 329   good
2056	acceptable 57   poor
Cluster 0	– very high-quality Cluster 1 – good
Cluster 2	– acceptable Cluster 3 - poor

Such expertise is the rising achievement of all faculties taken into consideration. The ratio of different comments is shown as below in Figure 3 for faculties in uncommon grades such as Lecturer, Associate Professor and each observation group as observed indicates a similar pattern towards each faculty category with the good and bad categories having almost the same baseline for both faculty categories.

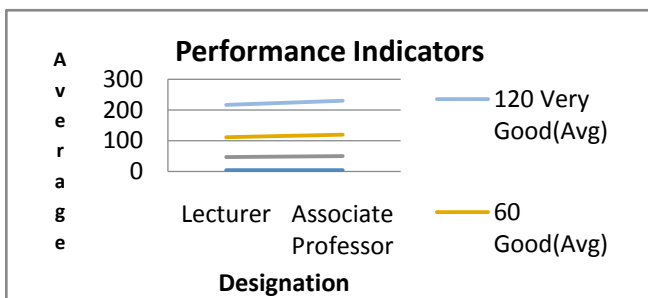


Fig.3 Comment Ratios versus Faculty Category

By running the grouping calculation on the arrangement information, the accurately characterized occasions were characterized, in this way producing the disarray framework that demonstrated the dispersion of the remark classification over the medium and furthermore gave the definite exactness regarding True Positive and False Positive rate by class. Next, the pre-handled preparing information was grouped and fixated on the most reduced and most elevated an incentive for every quality class; the input mean each worth was accounted for and appeared in the above chart. In like manner, Figure 3 mirrors the results of the bunch through perception gatherings and assignments. On the off chance that anticipated cycles were embraced, the pulling out of different sorts of assessment would have been very tedious or not doable.

After the text edit has been completed, the paper is ready for the template. Duplicate the template file by using the Save As command, and use the naming convention prescribed by your conference for the name of your paper. In this newly created file, highlight all of the contents and import your prepared text file. You are now ready to style your paper.

**4. Association Rule Analysis**

The student reaction data was subjected to association rule mining after the clustering was done. In order to disrupt the hidden patterns in the presentation of faculty and their conduct, the association rule mining was introduced. To figure out the best rules or relations, the Apriori Algorithm was used effectively. The minimum support range of the lower bound and upper bound was given between .1 and 1.0 and the algorithm was executed on 2890 properly classified instances.

**5. Results and Discussion**

In this report, the authors will show some results of the test data that will be used to compare calculations with manual calculations using the Rapid Miner software. The table below is the product of the test results. The highest value of the training aspects is 10.9 with the name of the Lecturer Initials, namely IDS, B, and AR, and the lowest value is 10.3 with the Initial Lecturers, namely NN, RRTK, WW, RRS, and EEM, based on Table 3 of the test data results. The highest score was 7.2 for the capability aspect, with the

initials of IIDS, BB, and AAR lecturers, with the lowest value of 6.8 with 2 initial names of lecturers, namely RRTK and EEM. The initial centroid is then randomly calculated, and the initial centroid is designated in Table 2 below.

**Table 2. Initial Centroids Data**

Centroids	X	Y	Z
I	10.67	17.79	7.11
II	10.53	17.55	7.02
III	10.47	17.46	6.98
IV	10.62	17.7	7.08

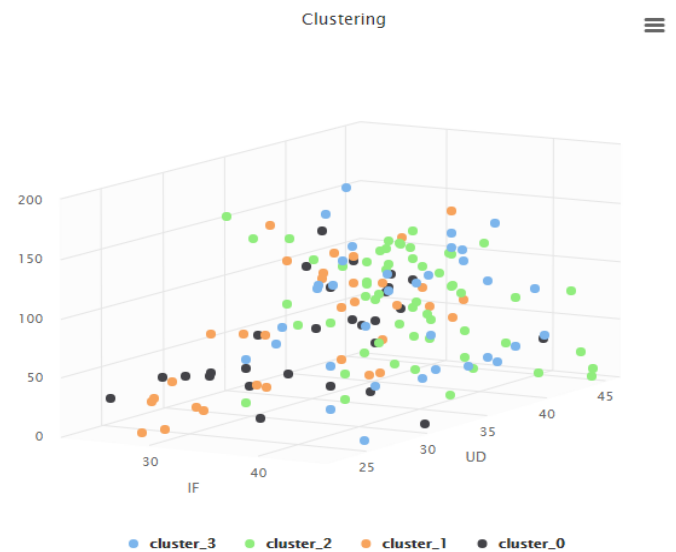
Name of Lecturer Initials	Preparation	Learning Process	Ability
IIF	10.56	17.61	7.04
UUD	10.53	17.55	7.02
SSA	10.62	17.7	7.08
FF	10.69	17.81	7.12
IIW	10.48	17.47	6.99

**Table 3 The results of the first iteration manual calculation between the distance of data and centroid**

Distance e1	Distance e2	Distance e3	Distance e4	c
0.22	0.07	0.38	0.12	2
0.99	0.00	0.32	0.18	2
0.11	0.18	0.30	0.00	4
0.03	0.32	0.44	0.14	1
0.39	0.10	0.02	0.28	3

Originally, the results of rapid miner computation version 9.0.3 were calculated by the aim centroid, i.e. cluster 0 = cluster 2, cluster 1 = cluster 3, cluster 2 = cluster 1, cluster 3 = cluster 4.

ID	Cluster	16311	16311	16311	16311	16311
1	Cluster_0	34	40	33	45	35
2	Cluster_0	33	45	35	45	44
3	Cluster_3	36	34	34	38	33
4	Cluster_2	32	31	34	35	44
5	Cluster_0	33	30	27	29	25



**Fig.4. Visualization of the results of calculations using rapid miners.**

Info: Information:  
 X-axis = planning factor  
 Y-axis = Method of learning  
 Axis of Z = capacity

**Conclusions**

It can be inferred that manual calculation has clustering results that are similar to the results of calculations using the rapid miner application from the results of a test comparing 2 calculations into the k-means clustering system to map aspect of lecturer performance indicators. There are distinct clustering results in the calculation using rapid miner version 9.0.3 because the centroid value in the rapid miner calculation is not calculated by the author but arbitrarily resolved by the rapid miner application.

**References**

[1].Hitendra Sarma T, Viswanath P and Eswara Reddy B 2012 A hybrid approach to speed-up the k-means clustering method.

International Journal of Machine Learning and Cybernetics(IJMLC), pages 1–11

- [2].Hambali, H. A., Abdullah, S., Lailee, S., Jamil, N., & Harun, H. (2016). Intelligent segmentation of fruit images using an integrated thresholding and adaptive K-means method (TsNKM). *Jurnal Teknologi*, 78(6-5), 13-20.
- [3].Amir Ahmad, Lipika De, “A K-Mean clustering algorithm for mixed numeric and categorical data” *Data & Knowledge Engineering Elsevier*,pp. 503-527,2007.
- [4].Nurzahputra, A., Muslim, M. A., & Khusniati, M. (2017). Penerapan Algoritma K-Means Untuk Clustering Penilaian Dosen Berdasarkan Indeks Kepuasan Mahasiswa. *Techno. Com*, 16(1), 17-24.
- [5].Purba, W., Tamba, S., & Saragih, J. (2018, April). The effect of mining data k-means clustering toward students profile model drop out potential. In *Journal of Physics: Conference Series* (Vol. 1007, No. 1, p. 012049). IOP Publishing.
- [6].Ramdhani, F., Hoyyi, A., & Mukid, M. A. (2015). Pengelompokan Provinsi di Indonesia berdasarkan Karakteristik Kesejahteraan Rakyat menggunakan Metode K-Means Cluster. *Jurnal Gaussian*, 4(4), 875-884.
- [7].Johan Oscar Ong. 2013. “Implementasi Algoritma K-Means Clustering Untuk Menentukan Strategi Marketing Presiden University.” Ed. *Jurnal Ilmiah Teknik Industri*, Vol.12, No.1.
- [8].Tutik Khotimah 2014. “Pengelompokan Surat Dalam Al Qur’an Menggunakan Algoritma K Means.” Ed. *Jurnal Simetris*, Vol 5 No 1.