

# Quantifying Performance Appraisal Parameters: A Forward Feature Selection Approach

Prafulla Bafna<sup>1\*</sup>, Samaya Pillai<sup>1</sup> and Dhanya Pramod<sup>2</sup>

<sup>1</sup>Symbiosis Institute of Computer Studies and Research (SICSR),  
Symbiosis International University (SIU), Atur Centre, Gokhale Cross Road, Model Colony,  
Pune – 411 016, Maharashtra, India; prafulla.bafna@sicsr.ac.in

<sup>2</sup>Symbiosis Centre for Information Technology (SCIT),  
Symbiosis International University (SIU), Plot 15, Phase 1, Hinjawadi,  
Pune - 411057, Maharashtra, India

## Abstract

**Objectives:** The objective of the paper is study and select optimal set of parameters present in performance appraisal (PA). It will result into the best PA. For achieving this, the technique used is “feature selection” and “clustering” and it is supported with the data analytical tool “R”. **Methods/Statistical Analysis:** The paper covers the data mining analysis. For achieving this, the technique used is “feature selection” and “clustering” and it is supported with the data analytical tool “R”. **Findings:** This paper focuses on performance appraisal, the multiple parameters available; to be exact it is 34 parameters. To select set of parameters, 13 from the set of 34, which when focused by the employees can have optimum PA. For achieving this, the technique used is “feature selection” and “clustering” and it is supported with the data analytical tool “R”. **Applications/Improvements:** Usable for every firm where employees have PA. In Organizations today, Human resource and performance appraisal has become very crucial. This is significant from the perception of both the management and the employees. The mechanism of measuring the performance appraisal has also evolved over a period. Recently there are multiple factors and parameters which are taken for measuring the performance appraisal of an individual employee. This complete process is very challenging for both, the employer and the employee. The employer i.e. the organization comes up with different mechanisms and keeps abreast with changing scenarios in order to be competitive in the industry. It is the employees who face the major difficulties in understanding and deciding what to address in their day to day work, which can get appreciation from the employers. Ultimately a good appraisal is every employee’s desire

**Keywords:** Clustering, Data Analytical Tool R, Feature Selection, Performance Appraisal

## 1. Introduction

In the current times, the task of performance management of the employees in organizations has undergone many developments from the earlier times<sup>25</sup>. Performance Appraisal has attained importance in last few decades. It is now observed as one of the most critical tools of HR<sup>3</sup>.

Performance Management has been an evolving domain. In the past organizations have been contemplative about this and overall development of the employees<sup>25</sup>. The work force and the organization have a combined effort for the completion of the desired work and success

of the organization within the industry. In today’s scenario the human resources have to be very committed for the organization and the organization needs to be committed for the employee and their overall manpower of the company<sup>5</sup>. The Performance appraisal has to be very structured and formal so as to implement it in a good manner for the organization.<sup>13</sup>

PA innovation as “appraisal intensification” is the tendency to apply traditional top-down PA (rater-ratee) various techniques and refinements whose aim is to enhance both the extent to which performance is under organizational surveillance and the differentiation of

\*Author for correspondence

employees into strong-weak, acceptable-unacceptable performers<sup>14</sup>. Two techniques are notable here:

(a) Multi-rater methods (like 360° appraisal)- which uses several data about an individual behavior, from peers, subordinates, etc.<sup>2,16</sup>

(b) Forced ranking method: here the PA system consists of managers applying percentage/quota of “high”, “average” and “poor performer” to populations of employees they appraise. This technique has been widely used by Ford, General Electric, Microsoft etc.<sup>7</sup>

In Indian Scenario, both the public and private sector require Performance Management and Appraisal, but there is a difference in their mechanism<sup>4</sup>. There are also a variety of different perspectives that are evident in the importance of performance management and its measurement<sup>40</sup>. Through rational performance measurement, Organization’s management can get the maximum reliance of employees.

## 1.1 Introduction to Feature Selection and Clustering

The method which is used to detect the relevant features is known Feature Selection (FS). It plays an important role in decision making process. For data driven decision support Data mining techniques are widely used. Before processing for data mining, it is good idea to carry out preprocessing. Feature selection techniques are generally used as a preprocessing technique. This technique minimizes attributes/dimensions. It eliminates irrelevant data. These noisy features decreases performance and efficiency of the algorithm which is used to run on the given dataset .The best set of features are selected from given dataset. In this paper we present a cluster based approach to select significant components that affect appraisal. Multiple factors are stated in appraisal form. It’s a need to choose only the significant parameters which affect the salary of employee. Expert knowledge is used to select the factors in the first step. The clusters are formed based hierarchical and partitioning family algorithms. Output of the first step is cluster purity. It is used as input to select the algorithm as a second step. This chosen algorithm is to be used in the next step. The forward FS technique is used to produce the clusters. Depending upon the cluster purity, optimal FS is achieved. Two step approach allows to eliminate unnecessary resources present in the next round of data selection. The data set used is related to manufacturing based company

## 2. Review of Literature

Use of computer, internet is resulted into production of huge amount of online and offline data<sup>18</sup>. Due to this extensive data, it may happen that the crucial parameters are being ignored during the process of decision making. It may in turn trigger anomalous reports.<sup>35</sup>

There are several popular techniques which are used to remove noisy attributes. E.g. (LDA), (PCA), (SVD). The techniques are targeted for Selecting the optimal set of features. (i.e. class label). Selected features can be evaluated using feature selection techniques. These are Fisher Score, Relief, etc.<sup>36,37</sup>

PA is the process used to determine performance of employee and communicates the information back to the employee. It also provides input for human resource planning (Riggio, 2003). An effectively designed, implemented appraisal system can produce an organization, manager and employee with a plethora of benefits<sup>14</sup>. There should be approach to minimizing subjective judgment in effective employee evaluation with many factors using hierarchical structure. To achieve hierarchical structure, clustering is one of the effective methods which leads to grouping according to the features of the objects.

Clustering is the process in which data elements whose properties are similar are placed in same group. So the items in different groups are dissimilar. The similarity measure can be used to evaluate clustering process. The process to find out this similarity between the items is known as cluster analysis<sup>41</sup>. The variety of applications is the reason to have different clustering techniques. It identifies outliers along with data distribution.<sup>21</sup>

To get accurate clusters several challenges should be considered. Selection of proper clustering approach is one of the issues. Validating clustering process is important to get reliable results. Cluster purity is used to validate the results. Irrelevant and unuseful data increases dimensions and affects the Purity of cluster. Choosing correct algorithm is one of the ways to solve this problem<sup>8</sup>. Cluster contents will be more correct Feature selection is followed by clustering techniques<sup>43</sup>. In some cases clusters won’t have any element because of the wrong choice of centroid. To define initial centroid is a big challenge.<sup>28</sup>

Keeping less important attributes which do not have any significance on final output leads to decrease the performance of algorithm. Researchers prefer to apply FS algorithm to improve efficiency. It is applied any on less data results into less cost and well performance. To group

the elements K- means Clustering algorithm is used . It is a unsupervised mining algorithm which uses a distance matrix to produce the clusters<sup>36,39</sup>. Feature selection is used with k-means clustering algorithm and used on dataset having 14 attributes in<sup>33</sup>. The researchers commented about the result obtained by applying FS techniques on clustering algorithm. There is 50% improvement in accuracy of the result, by applying FS techniques. But still there is need to have more efficient technique to develop the FS algorithm which will select optimal features and will produced highly accurate result<sup>6,9,32,34</sup>. Normally multiple search strategies available in literature. It is used with feature selection increases degree of optimality of final feature subset.

There are different metrics for evaluating clustering algorithms. Database indexing is an internal evaluation scheme, which uses quantities and attributes closer to the dataset. Cluster purity also can be measured used F-measure, entropy and precision. Entropy is used with various preprocessing methods such as wrapper, filter for feature elimination, reduction and selection<sup>12,23,24,29,38,42,44</sup>.

### 3. Research Design

#### 3.1 Data Collection and Pre-Processing

Collected dataset is from industrial sector. If the entire data set is given as an input it not only needs more time to run the experiment but also produces incorrect results. Decision making process can be fast, if representative and significant elements of data from given dataset is being chosen. Data consists of 366 X 34 dimensions and selected dimensions are 13. Base set attributes selected from domain knowledge.

#### 3.2 Selection of Clustering Algorithm

The first step of our approach is based on the survey that we carried out to determine the algorithm. It involves selection the algorithm which will be used for clustering. The selected algorithm is input for the second step that is feature selection. The comparative analysis of three algorithms based on experiments performed finalizes the algorithm. The data set is obtained by using expert's knowledge. It is a compact set of features. Algorithm producing maximum cluster purity is used as evaluation measure to decide algorithm for the next step of FS.

Results for three clustering algorithms are presented. Table 1 shows number of points present in each cluster. Table 2 indicates the different performance metrics to conclude the best algorithm. Lower the measure of Database Index, count of overlapped points and total withinness indicates improved performance of the algorithm.

**Table 1.** Cluster size Table

Algorithm	C1	C2	C3	C4
K-means	150	100	95	71
continuous K-means	100	100	95	121
Pairwise Agglomerative	50	200	70	96

**Table 2.** Cluster Validation Table

Algorithm	Database Index	Overlapping Points	Total With Innes
K-means	0.389	50	4256.3
continuous K-means	0.61	19	1432.5
Pairwise Agglomerative	0.812	92	1556.2

#### 3.3 Objectives

- To study the importance of feature selection.
- To apply feature selection method for calculation of optimal PA factors.
- To suggest optimal PA factors with their measure to each employee.

#### 3.4 Statistical Analysis

##### 3.4.1 Analysis and Discussion

###### 3.4.1.1 Selection of Optimal Feature Set

To avoid exponential Complexity of algorithm, present features are considered in combinations. Experts knowledge is used on increasing feature set. Continuous k-means is applied along with forward feature set. Table 3 and Table 4 show comparative analysis of algorithms.

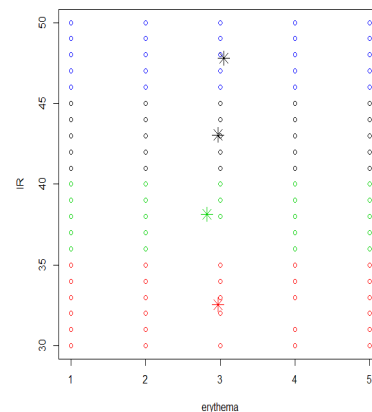
**Table 3.** Analysis of used algorithms

K-means algorithm. 1. overlapped clusters.	Continuous K-means algorithm non-overlapping clusters	Pairwise Agglomerative algorithm Different non overlapping clusters.
2. less changes in clusters of successive iterations	2. Prominent changes in the clusters of successive iterations.	2. In consecutive iterations important updates in the clusters are seen.
3. In the first step centroid of a clusters are random readings	3. Initial clusters are formed by visual observation. The nearest data points are considered.	3. Clusters are formed by using two points having the least distance.
4. Object in the cluster is input data points	4. Object in the cluster is input data points.	4. Object in the cluster is input data points
5. Slow execution	5. less time	5. Minimum twice steps

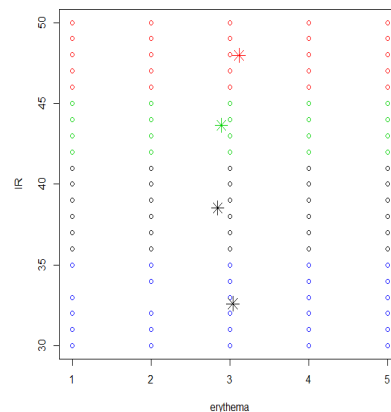
**Table 4.** Drawbacks of algorithms

K-means	Continuous K-means	Pairwise agglomerative
1. Suitable only for selected data.	----	1. Proper for few data elements.
2. Optimality is checked at every step, so produces accurate results.	----	2. Multiple steps takes long time
3. Before conducting an experiment k's value should be known.	----	3. less accurate results
	4. Before conducting an experiment k's value should be known.	4. Before conducting an experiment k's value should be known.

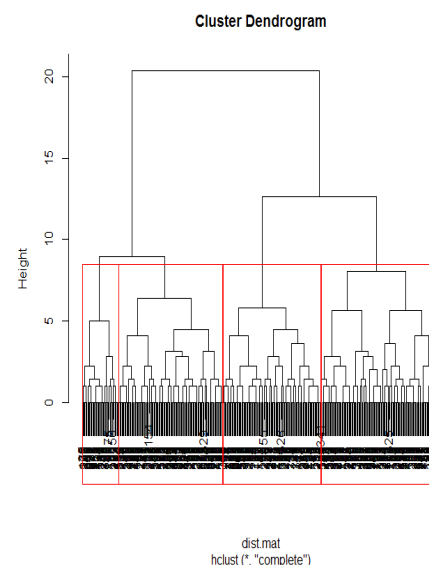
Graphical notations as A,B,C,D represent the clustering types wrt. number of parameters. Continuous K means and hierarchical agglomerative clustering algorithm is used on increased features gives best clusters .Best clusters are determined by the parameters such as error rate, Database index ,entropy, purity. Less entropy



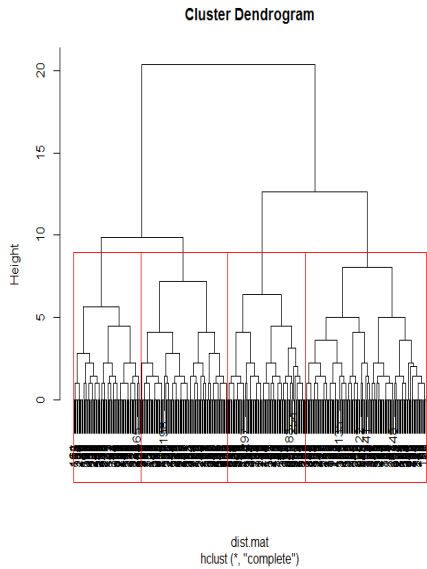
**Graph A.** Clusters' view for 367 X 34 data



**Graph B.** Clusters view for 367X 21 data



**Graph C.** Clusters' view for 367 X 34 data



**Graph D.** Clusters' view for 367X 21 data

and database index represent the good cluster quality<sup>14,26,41</sup>. The table shows optimum number of attributes. The study shows that 13 feature are selected by forward selection method from given features for further analysis. Table 5 shows the feature set. Table 6 shows the optimal feature number with cluster validating parameters.

**Table 5.** Features

Total feature set	Selected feature set
Name1,Name 2,date, location, department, salary, qualification, age, gender, innovations used , past knowledge used , resource utilization capacity,	competency, job knowledge, attitude towards work, purity, work performance, timeliness, learning skill 1, learning from training program, application of learning, knowledge development, team leading, attitude development, skilldevelopment.

## 4. Conclusion

Appraisal is one of the management's toughest and least satisfying tasks. And instead of thinking it as burden, it should be used to motivate and develop the employees. And also make a fair decision in deciding an appropriate salary and its progression path for the employees.

The proposed framework is applicable for a problem where clustering technique is used for decision making.

**Table 6.** Best Factors detection

No. of factors	Data Base index	SMSE (sum mean square error)	entropy	purity
10	0.8769	46.1783	0.6812	.42
14	0.4831	34.5623	0.6214	.7136
13	0.2468	8.26	0.0901	.8981
26	0.4231	28.4213	0.3981	.7621
30	0.4293	14.1342	0.2682	.7845
34	0.6541	31.3431	0.3216	0.5928

There are different data set repositories. These repositories are set of tables having multiple irrelevant attributes. Removal of useless attributes results into effective output. To achieve this reduced feature set are considered and then clustering algorithm is applied. After evaluating quality of the algorithm, it is finalized. It is then used for next step of FS. FS is applied before use of clustering algorithm results into removal noisy and less important attributes. Algorithm converges to solution fast and accurate. It minimizes the cost and efforts to collect the next round of data. So in the first step clustering algorithm is selected and then optimum features are selected. It increases the efficiency of decision process. In reference to the current study, this constituted 34 attributes related to the employee. These attributes were used during Appraisal evaluation. But many of them did not impact the Appraisal directly. Used the above said process, we could reduce the number of relevant attributes to 13. The study proves that the purity of the clustering based on these 13 parameters was far more better than the clustering achieved with 34 parameters.

## 5. Acknowledgement

A novel clustering approach to select optimal usability principles for educational websites, Prafulla Bafna, International Journal of Software and Web Sciences , ISSN (Print): 2279-0063, ISSN (Online): 2279-0071, Issue 11, December-2014 to February-2015 Sciences(IJSWS).

## 6. References

1. Bernardin HJ, Beatty R. Performance appraisal: Assessing human behavior at work. Kent, 1984.



2. Bracken D, Timmreck C, Fleenor J, Summers L. 360 Feedback From Another Angle. *Human Resource Management*. 2001; 40(1):3–20.
3. Bracken DW, Summers L, Fleenor J. High-tech 360. *Training and development*. 1998 Aug; 42–5.
4. Bhatia K, Jain P. A study of performance appraisal and organizational and organizational basis. A comparative study of BSNL and AIRTEL. *International Journal on Arts, Management and Humanities*. 2012; 1(1):63–72. ISSN No. (Online): 2319 – 5231 effectiveness in terms of individual
5. Mishra L. A Research Study on Employee Appraisal System Case of Hong Kong and Shanghai Banking Corporation (Hsbc Bank). *International Journal of Business and Management Invention*. 2013 Feb; 2(2):60–7. Available from: [www.ijbmi.org](http://www.ijbmi.org). ISSN (Online): 2319 – 8028, ISSN (Print): 2319 – 801X
6. Farahat AK. Data Mining (ICDM), An Efficient Greedy Method for Unsupervised Feature Selection, *IEEE*, 2005.
7. Gary L. The Controversial Practice of Forced Rating” in *Harvard Business Review*, 2001 Oct.
8. Doquire G, Verleysen M. Mutual information for feature selection with missing data computational Intelligence and Machine Learning, 2011.
9. Mamitsuka H. Principles of Data Mining and Knowledge Discovery. *Lecture Notes in Computer Science* springer, 2002.
10. Available from: <https://archive.ics.uci.edu/ml/datasets/Dermatology>. available on 01/03/2015
11. Liu H, Yu L. Toward Integrating Feature Selection Algorithms for Classification and Clustering.
12. Liu H. Evolving feature selection, intelligent systems, *IEEE*. 2005; 20(6):64–76.
13. Liu H. Evolving Feature Selection, *Intelligent Systems, IEEE*. 2005; 20(6).
14. Janosik SM, Creamer DG, Hirt JB, Winston RB, Saunders SA, Cooper DL. Supervising new professionals in student affairs: A guide for practioners. New York, NY: Brunner-Routledge, 2003.
15. Jenkins A. *Employment Relations in France : Evolution and Innovation*. New York, Kluwer Academic/Plenum Publishers, 2000.
16. Han J, Kamber M. *Data Mining: Concepts and Techniques*.
17. Jonson J, Olson A. Implementing Multiple Perspective Feedback : An Integrated Framework. *Human Resources Management Review*. 1996; 6(4):253–78.
18. Bijuraj LV. Clustering and its Applications. *Proceedings of National Conference on New Horizons in IT - NCNHIT*. 2013.
19. Witten LH, Frank Eibe. *Data miming: practical machine tools and techniques*
20. Latham GP, Skarlicki D, Irvine D, Siegel JP. In: Cooper CL, Robertson JT editors. London: John Wiley & Sons. 1993; 8(3):87–133
21. Lesh N. MERL, Zaki MJ. Scalable feature mining for sequential data, *Intelligent Systems and their Applications, IEEE*. 2000; 15(2).
22. Liu H, Torkkola K. Evolving feature selection ,*Intelligent Systems, IEEE*, 2005.
23. Ramaswami M, Bhaskaran R. A Study on Feature Selection Techniques in Educational Data Mining. *Journal of Computing*. 2009.
24. Mittal M, Sharma RK, Singh VP. Validation of K-means and Threshold based Clustering Method. *International Journal of Advancements in Technology*. 153–60.
25. Jashki MA et al. An iterative hybrid filter-wrapper approach to feature selection for document clustering. *Proceedings of the 22nd Canadian Conference on Advances in Artificial Intelligence 2009*; 5549. P. 74–85.
26. Murphy KR, Cleveland NJ. *Performance Appraisal: An organizational perspective*. Needham Heights, MA, Allyn & Bacon, 1991.
27. Murphy KR, Cleveland NJ. *Understanding performance appraisal – Social, organizational and goal based perspective*. Thousand Oaks: Sage Publications, 1995.
28. Murphy K. *Performance Measurement and Appraisal: Motivating Managers to Identify and Reward Performance*. in Bruns Ed. 1992; 37–62.
29. Agarwal P, Afshar Alam M, Biswas R. Issues, Challenges and Tools of Clustering Algorithms. *IJCSI International Journal of Computer Science Issues*. 2011 May; 8(3):2.
30. Bafna P, Gaikwad H. A hybrid approach to measure design improvement factor of website, *American International Journal of Research in Science, Technology, Engineering and Mathematics*. 2014 Dec-2015 Feb. (ISSN (Print): 2328-3491, ISSN (Online): 2328-3580, ISSN (CD-ROM): 2328-362).
31. Bafna P, Metkewar P, Shirwaikar S. Novel Clustering approach for Feature Selection. *American International Journal of Research in Science, Technology, Engineering and Mathematics*. 2014; 62–7.
32. Ratzburg WF. *Performance Appraisal*. 2009. Available from: <http://www.oocities.org/wilfratzburg/performance-appraisal.html>
33. Diao R, Shen Q. Feature Selection With Harmony Search. *IEEE Systems, Man, and Cybernetics*, 2012.
34. Diao R, Shen Q. Two New Approaches to Feature Selection with Harmony Search. *IEEE World Congress on Computational Intelligence*. 2010.
35. Ganda R, Chahar V. A Comparative Study on Feature Selection used Data Mining Tools. *International Journal*

- of Advanced Research in Computer Science and Software Engineering. 2013.
36. Roberts GE. Linkages between performance appraisal system effectiveness and rater and rate acceptance. *Review of Public Personnel Administration*. 1992; 12(3):19–41.
  37. Alelyani S, Tang J, Liu H. Feature Selection for Clustering: A Review. *public.asu.edu*. 2013.
  38. Alelyani S. Thesis on Feature Selection Stability: A Data Perspective, 2012.
  39. Saha S, Bandyopadya S. Some connectivity based cluster validity indices, *ACM used Soft Computing*. 2012 May; 12(5):1555–65.
  40. Beniwal S, Arora J. Classification and Feature Selection Techniques in Data Mining. *International Journal of Engineering Research and Technology (IJERT)*, Wadsworth Publishing, Boston. 2012 Aug; 1(6).
  41. Whitney JO. *The trust factor: Liberating profits and restoring corporate vitality*. New York: McGraw-Hill. 1994.
  42. He X, Ji M, Zhang C, Bao H. A Variance Minimization Criterion to Feature Selection Used Laplacian Regularization. *IEEE Transaction on Pattern Analysis and machine Intelligence*. 2011; 33(10).
  43. Li Z et al. Clustering-Guided Sparse Structural Learning for Unsupervised Feature Selection. *IEEE Transactions on Knowledge and Data Engineering (TKDE)*, 2014.
  44. Liu Z et al. A feature selection method for document clustering based on part-of-speech and word co- occurrence, *IEEE Conference on Fuzzy Systems and Knowledge Discovery*. 2010; 5:2331–4.