

Employee's Performance Analysis and Prediction using K-Means Clustering & Decision Tree Algorithm

SM Shamim¹

¹ Mawlana Bhashani Science and Technology University

Received: 11 December 2017 Accepted: 1 January 2018 Published: 15 January 2018

Abstract

Employee is the key element of the organization. The success or failure of an organization depends on the employee performance. Hybrid procedure based on Data Clustering and Decision Tree of Data mining method may be used by the authority to predict the employees' performance for the next year. This paper presents how data clustering method can be applied for evaluating the employee's performance as well in decision making process. Different performance evaluation factors like personality, punctuality, tact oral expression etc has been studied. The result of this paper predicts the number of employee those are selected for promotion or designation and discharged according to their performance. This study help to find out the inefficient employee, magnitude of inefficiency and aids to eliminate inefficiencies with a relatively easy to employ framework.

Index terms— data clustering, data mining, decision tree, employee's performance, prediction.

1 Introduction

Most of the organizations or companies have a formal performance evaluation system in which employee job performance is graded on a regular basis, usually once or twice a year. A good performance evaluation system can prominently benefit an organization. It helps employee behavior toward organizational aims by permitting employees know what is expected for them, and it yields information for making employment decisions, such as those regarding pay raises, promotion or releases. Developing and implementing an effective system is no easy task [1]. An Employee can improve their performance by way of monitoring the progression of their performance [2]. Machine learning algorithms i.e. clustering algorithm and decision tree of data mining technique can be used to find out the key characteristics of future prediction of an organization. Clustering is a method to group data into classes with identical characteristics in which the similarity of intra-class is maximized or minimized [3]. This method is most widely used procedure for future prediction.

The main objective of this paper is to partition Employees into homogeneous group according to their characteristics and abilities using clustering. This study makes use of cluster analysis to segment employees in to groups according to their performance. Decision tree has been used for making meaningful decision for the Employee. Based on the employee's performance results possible to take decision whether advanced training, talent enrichment or further qualification required or not. These applications also help administrative staff to enhance the quality of the organizations.

The structure of this paper is as follows. Section 2, the technical background described. In Section 3, Test Bed Setup has been explained. Section 4, evaluates the results after the experiment. Section 5, concludes the paper with future work.

2 II.

3 Background

Data Clustering is unsupervised and arithmetic data analysis procedure. Cluster analysis is used to segment a large set of data into subsets called clusters. Each cluster is a collection of data objects that are similar to one another place within the same cluster but are dissimilar to objects place other clusters. It is used to classify the same data into a homogeneous group. It's also used to operate on a large data-set to discover hidden pattern and relationship which helps to make decision quickly and efficiently. Data clustering and decision tree algorithm [4] has been used to evaluate the employee performance. Firstly, apply K-means clustering for separating Employees performance into four clusters which is Excellent, Good, Average and Poor according their Performance. Then apply Decision tree Algorithm for predicting next year Performance.

4 a) K-Means Clustering

K-means clustering [5] is a type of unsupervised learning, which is used in unlabeled data (i.e., data without defined categories or groups). The goal of this algorithm is to find groups in the data, with the number of groups represented by the variable K. The algorithm works iteratively to assign each data point to one of K groups based on the features that are provided. Data points are clustered based on feature similarity.

K-Means clustering intends to partition n objects into k clusters in which each object belongs to the cluster with the nearest mean. This method $J = \sum_{i=1}^n \sum_{j=1}^k ||x_{ij} - c_j||^2$

Where J is the object function, k number of cluster, n number of cases, and $||x_{ij} - c_j||^2$ is a chosen distance measure between a data point x_{ij} and the cluster center c_j , [6,7]. The algorithm and flow-chart of K-means clustering is given below.

Step 1: Accept the number of clusters to group data into and the dataset to cluster as input values

Step 2: Initialize the first K clusters

? Take first k instances or

5 ? Take Random sampling of k elements

Step 3: Calculate the arithmetic means of each cluster formed in the dataset.

Step 4: K-means assigns each record in the dataset to only one of the initial clusters.

Step 5: K-means re-assigns each record in the dataset to the most similar and re-calculates the arithmetic mean of all the clusters in the dataset. The flow chart of the k-means algorithm is given below.

6 Fig. 1: Flowchart of K-Means Clustering b) Decision Tree

Data mining [8] consists a set of techniques that can be used to extract relevant and interesting knowledge from data. It has several tasks such as association rule mining, classification and prediction and clustering. All classification methods are supervised learning techniques that classify data item into predefined class label. It is one of the most useful techniques in data mining to build classification models from an input data set. The used classification techniques commonly build prototypes that are used to guess future data developments. The basic algorithm for decision tree induction is a greedy algorithm that constructs decision trees in a top-down recursive divide-and-conquer manner [9].

Decision Tree generates a decision tree from the given training data. It is one of the most used techniques, since it creates the decision tree from the data given using simple equations depending mainly on calculation of the gain ratio, which gives automatically some sort of weights to attributes used, and the researcher can implicitly recognize the most effective attributes on the predicted target. As a result of this technique, a decision tree would be built with classification rules generated from it. Here Decision tree is used for predict Employee Performance. IV.

7 III.

8 Test Bed Setup

9 Result and Analysis

The main objective of this paper is to cluster/group employee according to their performance using K-means clustering and decision tree algorithm.

Four years data have been collected from an organization employee's database which consist 100 samples of data.

10 Fig . 2: Data without Clustering

By applying K-means clustering algorithm on the training data four group Excellent, Good, Medium and Poor has created according to employee's previous year performance. Figure ?? shows the initial values before clustering. One is excellent, second is good, third is average and the last one is poor. Figure ?? shows the 4 cluster such as Excellent, good, medium and poor based on their performance throughout 1 st year. Now the Employees four

94 group has been created. One is excellent, second is good, third is average and the last one is poor. Graphical
95 representation of these four categories is given below.

96 11 Range of

97 Using decision tree algorithms first calculate the entropy from which Information Gain (IG) has to estimate. There
98 are 4 attributes which are 1 st year, 2 nd year, 3 rd year and 4 th year Performance of Employee. Maximum
99 Information Gain attribute will be the root node. 1 st year attribute has highest Information Gain which is the
100 root node and the 4 th year attribute has the lowest Information Gain. Figure ?? shows the decision tree based
101 on some rules.

102 The following rules have been generated after calculating the entropy and information gain:

103 -If 1 st , 2 nd , 3 rd and 4 th year Performance is greater than 4.5 then the final result Excellent. -If 1 st year
104 Performance is greater than 4.5 but 2 nd , 3 rd and 4 th year Performance is greater than 4.0 and less than 4.5
105 then the final result will be good. -If 1st year and 2 nd year performance is greater than 4.0, but less than 4.5
106 and 3 rd year and 4 th year performance is greater than 3.0, but less than 4.0 then the final result Average. -If
107 1 st , 2 nd , 3 rd and 4 th year Performance is less than 3.0 then the final result will be poor.

108 From the generated rules administrative staff or authorities can easily predict he next year performance.
109 Moreover, from this prediction authority can take necessary steps for the qualified and inexpert Employees. The
110 authority may take the following action based on the performance -If the Employee is excellent then he/she will
111 get the opportunity of promotion. -If the Employee's Performance is good he/she need not to take special care
112 but involved themselves with some development activities.

113 -If the Employee's Performance is medium he/she need some training and try to understand the weak area of
114 him/her, then always prepared for improving themselves.

115 -If the Employee's Performance is Poor he/she will be selected for expulsion. Several rules have been created
116 based on four year performance of an organization's employee. If generated rules apply to predict 5 th year
117 performance based on previous years, prediction result will be nearly closed to the actual performance.

118 V.

119 12 Conclusions

120 This paper presents an overview of k-means clustering algorithm and decision tree algorithm. Kmeans clustering
121 algorithm is a common way to define classes of jobs. Here we apply K-means clustering algorithm for partitioning
122 Employee into different cluster based on their quality of Performance. We use decision tree algorithm for classify
123 Employee easily and take appropriate decision quickly. Several actions can be taken in this circumstance to
124 avoid any danger related to hiring poorly performed employee. Future work involves more proper data from
125 several companies. When the appropriate model is generated, these algorithms could be developed for predicting
126 performance of employees in any kind of organization. ^{1 2}

¹© 2018 Global Journals Employee's Performance Analysis and Prediction using K-Means Clustering & Decision Tree Algorithm

²Year 2 018 () C © 2018 Global Journals Employee's Performance Analysis and Prediction using K-Means Clustering & Decision Tree Algorithm

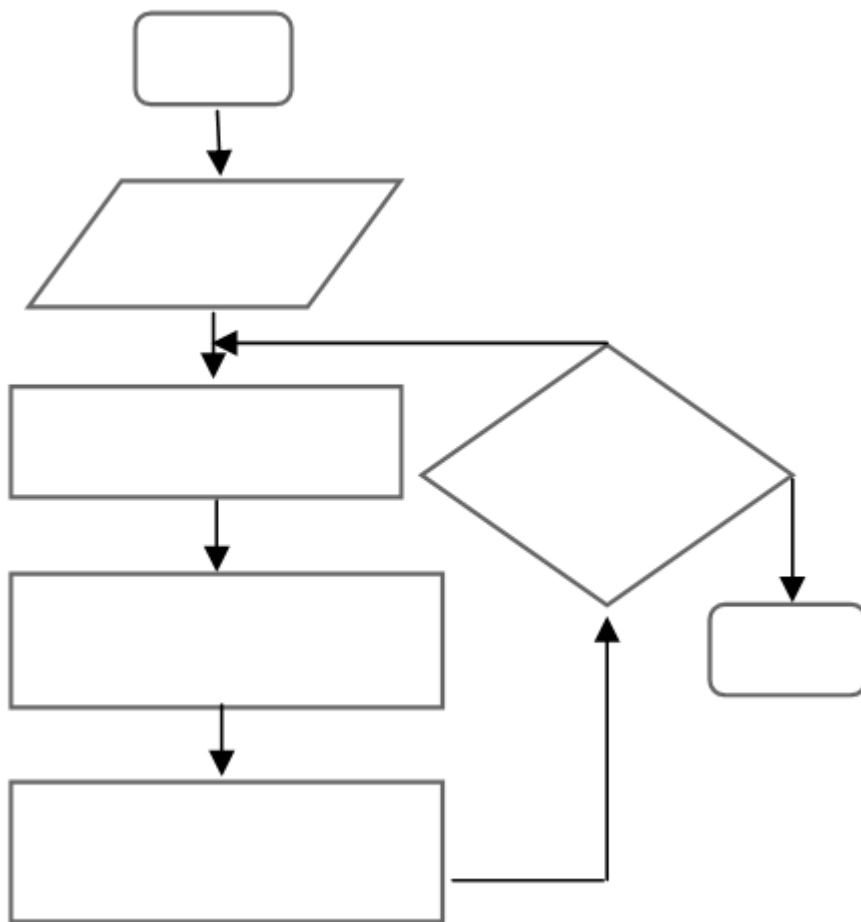
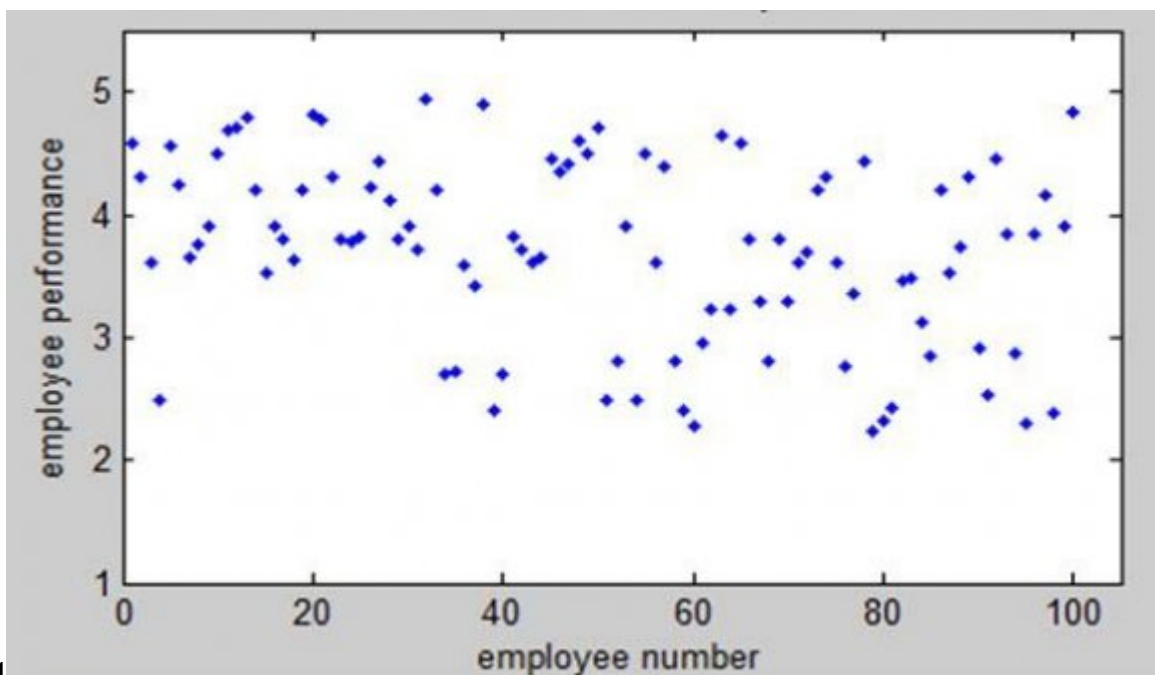


Figure 1:



34

Figure 2: Fig. 3 :Fig. 4 :

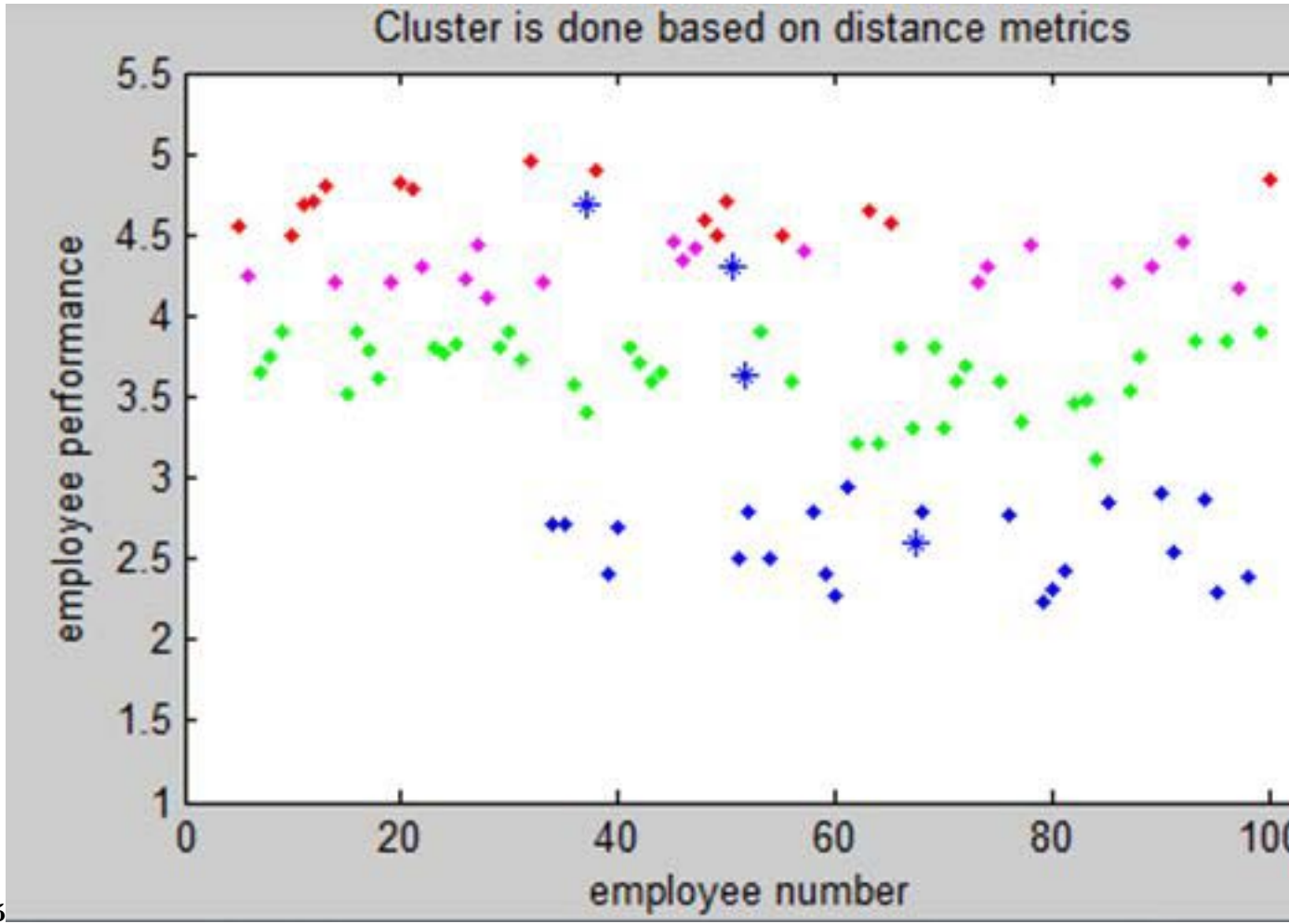


Figure 3: .Fig. 5 :

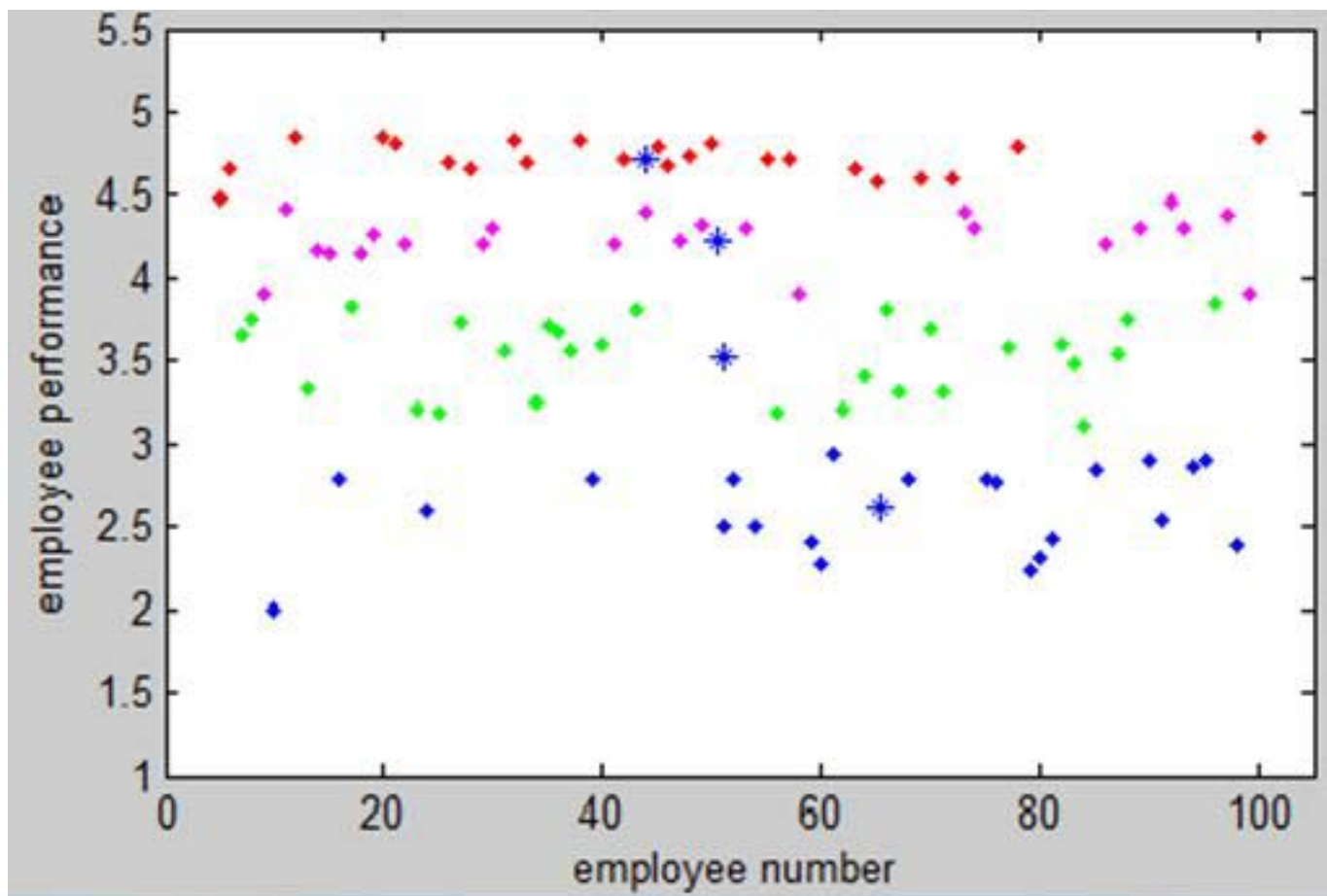


Figure 4:

1

Very good	5
Good	4
Average	3
Below Average	2
Poor	1

Figure 5: Table 1 :

2

Cluser	Category	No of Employee in different cluster				Percentage
		1 st yr	2 nd yr	3 rd yr	4 th yr	
Excellent	Quality \geq 4.5	17	23	20	22	20.50%
Good	4.0 \leq Quality $<$ 4.5	20	22	25	26	23.25%
Average	3.0 \leq Quality $<$ 4.0	40	32	34	36	35.50%
Poor	Quality $<$ 3.0	23	23	21	16	20.75%

Figure 6: Table 2 :

-
- 127 [International Journal of Advanced Computer Science and Applications ()] , *International Journal of Advanced*
128 *Computer Science and Applications* 2012. 3 (2) .
- 129 [Azar et al. ()] ‘A Model for Personnel Selection with A Data Mining Approach: A Case Study in A Commercial
130 Bank: Original Research’. A Azar , M V Sebt , P Ahmadi , A Rajaeian . *SA Journal of Human Resource*
131 *Management* 2013. 11 (1) p. .
- 132 [Lakshmi et al. ()] ‘An Analysis on Performance of Decision Tree Algorithms using Student’s Qualitative Data’.
133 T M Lakshmi , A Martin , R M Begum , V P Venkatesan . *International Journal of Modern Education and*
134 *Computer Science* 2013. 5 (5) p. 18.
- 135 [Shovon et al. ()] *An Approach of Improving Students Academic Performance by using k Means Clustering*
136 *Algorithm and Decision Tree*, M Shovon , H Islam , M Haque . arXiv:1211.6340. 2012. (arXiv preprint)
- 137 [Oyelade et al. ()] *Application of k Means Clustering algorithm for prediction of Students Academic Performance*,
138 O J Oyelade , O O Oladipupo , I C Obagbuwa . arXiv:1002.2425. 2010. (arXiv preprint)
- 139 [Varghese et al.] *Clustering Student Data Fig. 6: Decision Tree Based on Rules*, B M Varghese , A Unnikrishnan
140 , G Scientist , N P O L Kochi , Kochi .
- 141 [Priyama et al. ()] ‘Comparative analysis of decision tree classification algorithms’. A Priyama , R G Abhijeeta
142 , A Ratheeb , S Srivastavab . *International Journal of Current Engineering and Technology* 2013. 3 (2) p. .
- 143 [Hameed and waheed ()] ‘Employee Development and it’s Affect on Employee Performance a Conceptual
144 Framework’. A Hameed , A &waheed . *International journal of business and social science* 2011. 2 (13)
145 .
- 146 [Nazeer and Sebastian ()] ‘Improving the Accuracy and Efficiency of the k-means Clustering Algorithm’. K A
147 Nazeer , M P Sebastian . *Proceedings of the World Congress on Engineering, (the World Congress on*
148 *Engineering)* 2009. 1 p. .
- 149 [to Characterize Performance Patterns Int. J. Adv. Comput. Sci. Appl ()] ‘to Characterize Performance Pat-
150 terns’. *Int. J. Adv. Comput. Sci. Appl* 2010. p. .
- 151 [Al-Radaideh and Nagi] *Using Data Mining Techniques to Build a Classification Model for Predicting Employees*
152 *Performance*, Q A Al-Radaideh , E Nagi .