

## Towards applying Data Mining Techniques for Talent Mangement

Hamidah Jantan<sup>1</sup>, Abdul Razak Hamdan<sup>2</sup>, Zulaiha Ali Othman<sup>2</sup>

<sup>1</sup>Faculty of Computer Science and Mathematics, UiTM Terengganu  
23000, Dungun, Terengganu, Malaysia  
e-mail: hamidahjtn@tganu.uitm.edu.my

<sup>2</sup>Faculty of Information Science and Technology  
National University of Malaysia (UKM)  
43600 Bangi, Selangor  
e-mail: arh,zao@ftsm.ukm.my

**Abstract.** Among the challenges of HR professionals is to manage organization's talents, especially to ensure the right person assign to the right job at the right time. This paper presents an overview of some of the talent management problems that can be solved by using Data mining techniques. Besides that, in this study, we attempt to implement one of the talent management tasks i.e. identifying potential talent by predicting their performance using past experience knowledge. An example demonstrates the feasibility of the suggested Data Mining techniques for the employee performance data. Finally, this study proposes the potential Data Mining Techniques for talent forecasting.

**Keywords:** Data Mining; Human Resource Management; Talent Management; Talent forecasting; Performance prediction.

### 1. Introduction

Human Resource (HR) applications that are embedded with Artificial Intelligent (AI) techniques can be used to help decision makers to solve unstructured decisions. In the advancement of AI technology, there are many techniques can be used to advance the capabilities of HR application. Data Mining is one of AI technology that has been developed for exploration and analysis in large quantities of data to discover meaningful patterns and rules. In actual fact, such data in HR data can provide a rich resource for knowledge discovery and decision support tools. So far, the techniques and application of Data Mining have not attracted much attention in Human Resource Management (HRM) field [1]. In this study, we attempt to use this approach in managing talent i.e. to identify existing talent by predicting their performance using the past experience knowledge.

Basically, HRM is a comprehensive set of managerial activities and tasks concerned with developing and maintaining a workforce-human resource. HRM aims to facilitate organizational competitiveness; enhance productivity and quality; promote individual growth and development; and complying with legal and social obligation [2]. Besides that, in any organizations, they need to struggle effectively in term of cost, quality, service or innovation. All these depend on having enough right people, with the right skills, deployed in the appropriate locations at appropriate points in time. Recently, among the challenges of HR professionals is managing talent, especially to ensure the right person for the right job at the right time. These tasks involve a lot of managerial decisions, which are sometimes very uncertain and difficult. In reality, current HR decision practices depend on various factors such as human experience, knowledge, preference and judgment. These factors can cause inconsistent, inaccurate, inequality and unexpected decisions. As a result, especially in promoting individual growth and development, this situation can often make people sense injustice. Besides that,

in future, this can influence the organization productivity. Identifying the existing talent is one of the top talent management challenges [3], and to handle this issue, it can be done by predicting the talent based on their performance using Data Mining technique. For that reason, this study aims to suggest the potential techniques in Data Mining for talent forecasting. Those potential techniques can be embedded with HR decision support applications in order to produce a system that can help managers decide the right employee for the right job. This paper is organized as follows. The second section describes the related works on Data Mining in HR and talent management concept. The third section discusses the potential Data Mining techniques that can be used in talent forecasting. Finally, the paper ends with Section 4 where the concluding remarks and future research directions are identified.

## 2. Related Works

### 2.1. Data Mining

Data mining (DM) is an approach currently receiving great attention and is being recognized as a newly emerging analysis tool [4]. Recently, Data mining has given a great deal of concern and attention in the information industry and in society as a whole. This is due to the wide accessibility of enormous amounts of data and the important need for turning such data into useful information and knowledge [5]. Computer application such as Decision Support System (DSS) interfaces with DM tool can help executives to make more informative and objectives decisions. Besides that, it can help managers to retrieve, summarize and analyze decision related data to make wiser and more informed decisions. Data mining problems are generally categorized as clustering, association, classification and prediction [1, 6]. Over the years, the Data mining has involved various techniques including statistics, neural network, decision tree, genetic algorithm, and visualization techniques. Besides that, Data mining has been applied in many fields such as finance, marketing, manufacturing, health care, customer relationship and etc. Nevertheless, its application in HRM is rare [6].

### 2.2. Data Mining in HR Applications

Nowadays, there are some researches on solving HRM problems that uses Data mining approach[1]. Table 1 lists some of the HR applications that use Data mining, and it shows there are few discussions on prediction applications use that technique in human resource domain. This technique is usually used in personnel selection to choose the right candidates for a job.

TABLE 1  
DATA MINING IN HR APPLICATIONS

Data Mining method used	Activity in HRM
<i>Fuzzy Data Mining and Fuzzy Artificial Neural Network</i>	Employee development – Project Assignment [7]
<i>Decision tree</i>	Personnel selection [6], Job attitudes [8]
<i>Association rule mining</i>	Employee Development – Training [9]
<i>Rough Set Theory</i>	Personnel Selection – Recruit and Retain Talents [10]
<i>Fuzzy Data Mining</i>	Personnel Selection [11]

Prediction applications in HRM are infrequent, there are some examples such as to predict the length of service, sales premiums, to persistence indices of insurance agents and analyze mis-operation behaviors of operators[6]. For that reasons, in this study, we attempts to use Data mining techniques to forecast potential employees as a part of talent management task.

### 2.3. Talent Management

In any organization, talent management is becoming an increasingly crucial method of approaching HR functions. Talent is considered as any individual who has the capability to make a significant difference to the current and future performance of the organization [12]. In fact, managing talent involves human resource planning that regards processes for managing people in organization. Besides that, talent management can be

defined as an outcome to ensure the right person is in the right job; process to ensure leadership continuity in key positions and encourage individual advancement; and decision to manage supply, demand and flow of talent through human capital engine [13]. In HRM, talent management is very important and need some attentions from HR professionals. There are some issues related to talent management shown in Fig. 1.

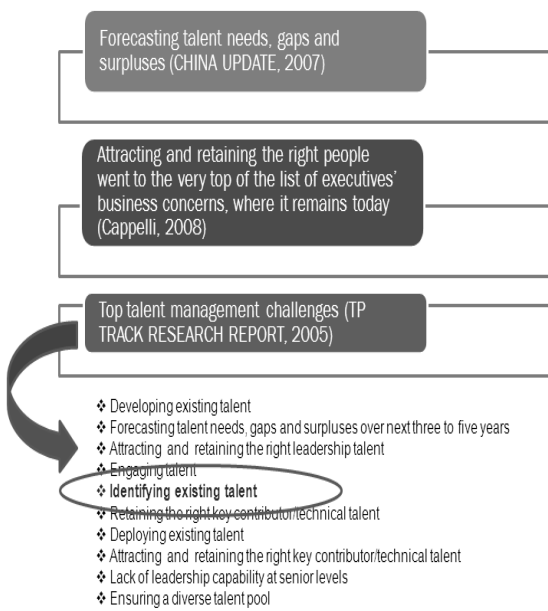


Fig. 1 : Talent Management Challenges

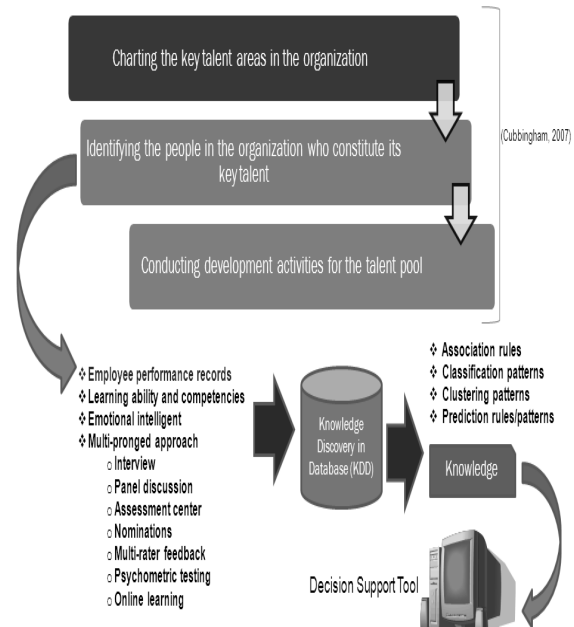


Fig. 2 : Data Mining for Talent Management

Besides that, Fig.1 also shows the TP Track Research Report that found among the top current and future talent management challenges are developing existing talent; forecasting talent needs; attracting and retaining the right leadership talent; engaging talent; identifying existing talent; attracting and retaining the right leadership and key contributor; deploying existing talent; lack of leadership capability at senior levels and ensuring a diverse talent pool [3]. In this study, we focus on one of the talent management challenge that is to identify the existing talent regarding to the key talent in an organization by predicting their performance. In this case, we use the past data stored in the existing databases. The talent management process consists of recognizing the key talent areas in organization, identifying the people in the organization who constitute its key talent, and conducting development activities for the talent pool to retain and engage them and have them ready to move into more significant roles [13] is illustrated in Fig. 2. These processes involve HR activities that need to be integrated into an effective system [14].

## 2.4. Data Mining for Talent Management

Basically, most of the Data Mining researches in HR problems domain focus on personnel selection task and few apply in other activities such as planning, training, talent management and etc. Recently, with the new demands and the increased visibility of HR management, thus, HRM seeks a strategic role by turning to Data Mining methods [1]. This can be done by identifying generated patterns from the existing data in HR databases as useful knowledge. In this study, we focus on identifying the patterns that relate to the talent. The patterns can be generated by using some of the major Data mining techniques and it is shown in Fig. 3. The matching of Data mining problems and talent management needs are very important, in a way to determine the suitable Data Mining techniques.

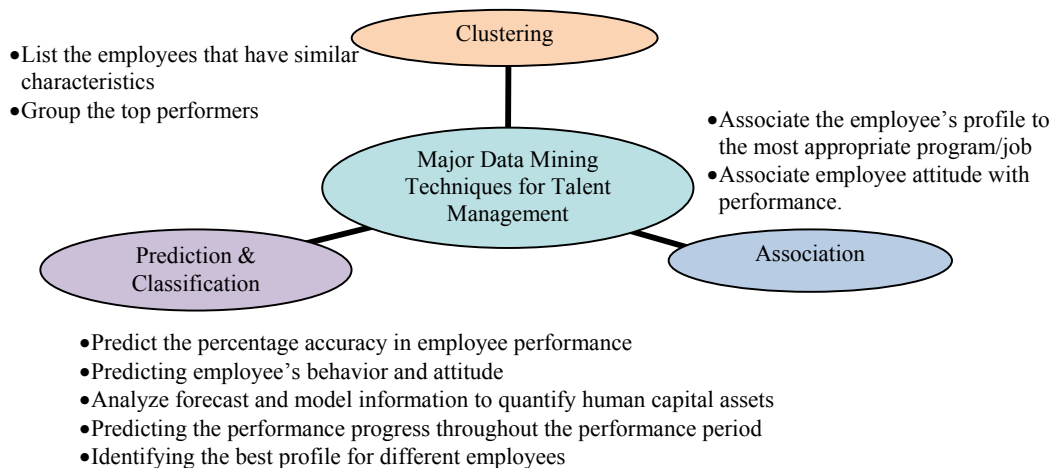


Fig. 3 : Data Mining Problems for Talent Management

### 3.0 Potential Data Mining Techniques for Talent Forecasting

Data mining technique is the best linear unbiased estimator, decision tree and neural network and is found useful in developing predictive models in many fields [4]. Table 2 lists some of the characteristics for neural network and decision tree techniques.

TABLE 2  
POTENTIAL DATA MINING TECHNIQUES

Data Mining Techniques	Characteristics	Techniques
<i>Decision Tree</i>	1. Usually used for classification and prediction tasks. 2. Produces a model which may represent interpretable rules or logic statement and more suitable for predicting categorical outcomes 3. Non-parametric: suited to capturing an functional form relating independent and dependent variables 4. Easy to interpret, computationally inexpensive and capable to dealing with noisy data. Model prediction is explainable to the model user. 5. Automatic interaction detection – find quickly significant high-order interactions 6. More informative outputs	<ul style="list-style-type: none"> <li>• C4.5</li> <li>• Random forest</li> </ul>
<i>Neural Network</i>	Provide a variety of powerful tools for optimization, function approximation, pattern classification and modeling	<ul style="list-style-type: none"> <li>• Multi Layer Perceptron</li> <li>• Radial Basic Function Network</li> </ul>

In this study, we attempts to use classifier algorithm C4.5 and Random Forest for decision tree; and Multilayer Perceptron(MLP) and Radial Basic Function Network for neural network. In the initial stage of this study, we run the selected classifier algorithms for the sample of employee data. In this case, we focus on the accuracy of the techniques to find the suitable classifier for HR data. The employee data contains 53 related attributes from the five evaluation performance factors that shown in table 3. The data are for professional and management employees from higher education institution.

TABLE 3  
NUMBER OF ATTRIBUTES FOR EACH FACTOR

Factor	Attributes
Background	7
Previous Performance	15
Knowledge and expertise	17
Management skill	8
Personal characteristics	6

The experiment of each algorithm is for 10 random processes for each selected training and test models. The results on the accuracy for the selected classifier algorithms are shown in table 4-7.

TABLE 4  
THE ACCURACY OF C4.5 FOR SELECTED MODEL

Model	Accuracy										
Train:Test	Rand1	Rand2	Rand3	Rand4	Rand5	Rand6	Rand7	Rand8	Rand9	Rand10	Average
90:10	60.00	90.00	90.00	100.00	80.00	100.00	70.00	70.00	90.00	40.00	79.00
80:20	75.00	95.00	90.00	95.00	90.00	90.00	85.00	80.00	65.00	95.00	86.00
70:30	96.67	96.67	83.33	86.67	73.33	96.67	73.33	76.67	73.33	86.67	84.33
60:40	75.00	77.50	75.00	72.50	95.00	95.00	77.50	85.00	85.00	77.50	81.50
50:50	86.00	84.00	74.00	98.00	88.00	84.00	80.00	54.00	78.00	98.00	82.40
40:60	83.33	86.67	80.00	68.33	63.33	73.33	71.67	75.00	63.33	80.00	74.50
30:70	80.00	68.57	84.29	74.29	78.57	58.57	74.29	68.57	81.43	61.43	73.00
20:80	71.25	58.75	56.25	61.25	55.00	46.25	66.25	53.75	71.25	73.75	61.38
10:90	52.22	62.22	61.11	46.67	52.22	53.33	65.56	50.00	43.33	34.44	52.11

TABLE 5  
THE ACCURACY OF RANDOM FOREST FOR SELECTED MODEL

Model	Accuracy										
Train:Test	Rand1	Rand2	Rand3	Rand4	Rand5	Rand6	Rand7	Rand8	Rand9	Rand10	Average
90:10	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
80:20	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
70:30	100.00	100.00	100.00	100.00	96.67	100.00	100.00	100.00	96.67	100.00	99.33
60:40	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
50:50	96.00	100.00	98.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	99.40
40:60	100.00	96.67	100.00	100.00	96.67	100.00	96.67	100.00	96.67	100.00	98.67
30:70	97.14	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	99.71
20:80	100.00	72.50	68.75	100.00	100.00	75.00	100.00	76.25	100.00	100.00	89.25
10:90	72.22	80.00	47.78	45.56	71.11	83.33	73.33	76.67	74.44	74.44	69.89

TABLE 6  
THE ACCURACY OF MLP FOR SELECTED MODEL

Model	Accuracy										
Train:Test	Rand1	Rand2	Rand3	Rand4	Rand5	Rand6	Rand7	Rand8	Rand9	Rand10	Average
90:10	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
80:20	100.00	100.00	100.00	100.00	95.00	100.00	100.00	100.00	100.00	100.00	99.50
70:30	96.67	96.67	96.67	96.67	93.33	96.67	96.67	96.67	96.67	96.67	96.34
60:40	97.50	95.00	95.00	97.50	95.00	97.50	95.00	97.50	95.00	87.50	95.25
50:50	90.00	90.00	88.00	94.00	92.00	94.00	94.00	82.00	92.00	100.00	91.60
40:60	91.67	95.00	96.67	88.33	83.33	93.33	85.00	86.67	86.67	83.33	89.00
30:70	77.14	82.86	87.14	84.29	71.43	82.86	90.00	84.29	85.71	80.00	82.57
20:80	73.75	66.25	61.25	73.75	73.75	58.75	75.00	72.50	87.50	76.25	71.88
10:90	60.00	65.56	50.00	53.33	47.78	61.11	71.11	62.22	56.67	55.56	58.33

TABLE 7  
THE ACCURACY OF RADIAL BASIC FUNCTION NETWORK FOR SELECTED MODEL

Model	Accuracy										
Train:Test	Rand1	Rand2	Rand3	Rand4	Rand5	Rand6	Rand7	Rand8	Rand9	Rand10	Average
90:10	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
80:20	100.00	100.00	100.00	100.00	95.00	100.00	100.00	100.00	100.00	100.00	99.50
70:30	100.00	100.00	96.67	96.67	96.67	100.00	100.00	100.00	96.67	96.67	98.34
60:40	97.50	100.00	100.00	100.00	97.50	100.00	100.00	100.00	97.50	97.50	99.00
50:50	98.00	100.00	98.00	100.00	98.00	98.00	100.00	98.00	98.00	100.00	98.80
40:60	98.33	98.33	98.33	100.00	98.33	100.00	98.33	100.00	98.33	100.00	99.00
30:70	77.14	85.71	100.00	97.14	92.86	91.43	98.57	82.86	98.57	97.14	92.14
20:80	98.75	71.25	68.75	97.50	68.75	68.75	95.00	68.75	96.25	87.50	82.13
10:90	58.90	64.44	54.44	55.56	56.67	66.67	54.44	58.89	43.33	27.78	54.11

The accuracy for each of the classifier algorithm is shown in Table 8. This initial experiments show that the accuracy of this two data mining techniques are higher than 70% and it can be considered as indicators for the suitable Data Mining techniques for HR data. As we mentioned before, the purpose of this experiment is to find out the accuracy of the popular prediction techniques in Data mining. In order to find the best Data Mining techniques, other Data Mining techniques such as SVM (support vector machine), Rough Set Theory, Fuzzy clustering and many others can be added as other classifier algorithm to support the accuracy of the these techniques. The results show us the suitability to use decision tree or neural network for talent forecasting.

TABLE 8  
THE ACCURACY FOR EACH CLASSIFIER

Classifier Algorithm	Accuracy
C4.5/J48	95.14%
Random forest	74.91%
Multilayer Perceptron (MLP)	87.16%
Radial basis function network	91.45%

## 4.0 Conclusion

This paper has described the significance of the study, background of data mining, data mining in human resource application and an overview of talent management. From the literature study, most researchers have

discussed HR applications from different type of application. However, there should be more HR applications and Data Mining techniques applied to different problem domains in HRM field research in order to broaden our horizon of academic and practice work on HR applications using Data Mining techniques. Due to these reasons, we propose the suitable Data Mining techniques for performance prediction based on initial experiment. For future work, the data in HR can be tested using other Data Mining techniques to find out the best accuracy of the techniques, especially for talent management data. Besides that, the relevant of attributes should be considered as a factor to the accuracy of the classifier. In next experiment, attribute reduction experiment should be take place in order to choose the relevant attributes for each of the factor. Once the relevant attributes are attained, the next modeling steps can be established to recommend. Finally, the ability to continuously change and obtain new understanding is the power of HR application, and this can be the HR applications of future work.

## 5.0 References

- [1] Ranjan, J., "Data Mining Techniques for better decisions in Human Resource Management Systems". International Journal of Business Information Systems, 2008. **3**(5): p. 464-481.
- [2] DeNisi, A.S. and R.W. Griffin, "Human Resource Management". 2005, New York: Houghton Mifflin Company.
- [3] A TP Track Research Report Talent Management: "A State of the Art". 2005, Tower Perrin HR Services.
- [4] Tso, G.K.F. and K.K.W. Yau, "Predicting electricity energy consumption : A comparison of regression analysis, decision tree and neural networks". Energy, 2007. **32**: p. 1761 - 1768.
- [5] Han, J. and M. Kamber, "Data Mining : Concepts and Techniques". 2006, San Francisco: Morgan Kaufmann Publisher.
- [6] Chien, C.F. and L.F. Chen, "Data mining to improve personnel selection and enhance human capital: A case study in high-technology industry". Expert Systems and Applications, 2008. **34**(1): p. 380-290.
- [7] Huang, M.J., Y.L. Tsou, and S.C. Lee, "Integrating fuzzy data mining and fuzzy artificial neural networks for discovering implicit knowledge". Knowledge-Based Systems, 2006. **19**(6): p. 396-403.
- [8] Tung, K.Y., et al., "Mining the Generation Xer's job attitudes by artificial neural network and decision tree - empirical evidence in Taiwan". Expert Systems and Applications, 2005. **29**(4): p. 783-794.
- [9] Chen, K.K., et al., "Constructing a Web-based Employee Training Expert System with Data Mining Approach", in Paper in The 9th IEEE International Conference on E-Commerce Technology and The 4th IEEE International Conference on Enterprise Computing, E-Commerce and E-Services (CEC-EEE 2007). 2007.
- [10] Chien, C.F. and L.F. Chen, "Using Rough Set Theory to Recruit and Retain High-Potential Talents for Semiconductor Manufacturing". IEEE Transactions on Semiconductor Manufacturing, 2007. **20**(4): p. 528-541.
- [11] Tai, W.S. and C.C. Hsu (2005) "A Realistic Personnel Selection Tool Based on Fuzzy Data Mining Method". [http://www.atlantis-press.com/php/download\\_papaer?id=46](http://www.atlantis-press.com/php/download_papaer?id=46).
- [12] Lynne, M., "Talent Management Value Imperatives : Strategies for Execution". 2005, The Conference Board.
- [13] Cubbingham, I., "Talent Management : Making it real". Development and Learning in Organizations, 2007. **21**(2): p. 4-6.
- [14] CHINA UPDATE (2007) "HR News for Your Organization : The Tower Perrin Asia Talent Management Study". <http://www.towersperrin.com>.