

Human Talent Prediction in HRM using C4.5 Classification Algorithm

Hamidah Jantan

Faculty of Computer and Mathematical Sciences
Universiti Teknologi MARA (UiTM) Terengganu,
23000 Dungun, Terengganu, Malaysia
Email: hamidahjtn@tganu.uitm.edu.my

Abdul Razak Hamdan and Zulaiha Ali Othman

Faculty of Information Science and Technology
Universiti Kebangsaan Malaysia (UKM)
43600 Bangi, Selangor, Malaysia
Email: {arh,zao}@fism.ukm.my

Abstract—In HRM, among the challenges for HR professionals is to manage an organization's talents, especially to ensure the right person for the right job at the right time. Human talent prediction is an alternative to handle this issue. Due to that reason, classification and prediction in data mining which is commonly used in many areas can also be implemented to human talent. There are many classification techniques in data mining techniques such as Decision Tree, Neural Network, Rough Set Theory, Bayesian theory and Fuzzy logic. Decision tree is among the popular classification techniques, which can produce the interpretable rules or logic statement. The generated rules from the selected technique can be used for future prediction. In this article, we present the study on how the potential human talent can be predicted using a decision tree classifier. By using this technique, the pattern of talent performance can be identified through the classification process. In that case, the hidden and valuable knowledge discovered in the related databases will be summarized in the decision tree structure. In this study, we use decision tree C4.5 classification algorithm to generate the classification rules for human talent performance records. Finally, the generated rules are evaluated using the unseen data in order to estimate the accuracy of the prediction result.

Keywords-Human talent; Classification, Prediction; Decision Tree; C4.5 classifier; Classification Algorithm

I. INTRODUCTION

¹Nowadays, there are many areas, such as in finance, medical, marketing, stock market, telecommunication, manufacturing, health care and customer relationship, which have adapted data mining techniques [1-4] [5-10] [11-13] [14-16]. However, the used of classification techniques in data mining approach does not attract much attention among people in Human Resource (HR) field [17, 18]. HR data provides a rich resource for knowledge discovery for decision support system development. In addition, today's organization has to struggle effectively in terms of cost, quality, service or innovation. The success of these tasks depends on having enough right people with the right skills, deployed in the appropriate locations at the appropriate point

¹ This research was conducted as a part of the science fund project funded by MOSTI (Ministry of Science, Technology and Innovation), Malaysia (01-01-01-SF0236).

of time which is known as talent management. Managing an organization talent has become one of the challenges to the HR professionals. This task involves a lot of managerial decisions in order to decide the right person for the right job at the right time. Sometimes, these types of decisions are very uncertain and difficult; and it depends on various factors such as human experience, knowledge, preference and judgment. The process to identify an existing talent in organization is among the top talent management challenges and becomes a never ending issue [19].

In data mining tasks, classification and prediction is among the popular task for knowledge discovery and future plan. The classification process is known as supervised learning, where the class level or classification target is already known. There are many techniques used for classification in data mining such as Decision Tree, Bayesian, Fuzzy Logic, Support Vector Machine (SVM), Artificial Immune System (AIS), Neural Network, Rough Set Theory, Genetic Algorithm and Nearest Neighbor. Decision tree is among the powerful classification algorithms as stated in some studies [20-23]. The decision tree technique has its advantages such as it can produce a model which may represent interpretable rules or logic statement; it is more suitable for analyzing categorical outcomes; it is non-parametric which is suited to capture a functional form relating independent and dependent variables; easy to interpret, computationally inexpensive, capable in dealing with noisy data, its prediction model is explainable to the user, it has automatic interaction detection to find the significant high-order interactions quickly, and it can produce more informative outputs [18, 20]. In fact, there are many techniques from the decision tree family such as C4.5, NBTree, SimpleCart, REPTree, BFTree and others. The C4.5 classification algorithm is easy to understand as the derived rules have a very straightforward interpretation. Due to these reasons, this study is aimed to use this classification algorithm to handle issue on human talent prediction.

In this study, recommendation for promotion (yes/no) is considered as the target class in classification process. For human talent dataset, we used employees' data from one of Malaysian higher learning institutions as training dataset. The first phase of mining process, the training dataset is prepared using the data mining preprocessing task. In the second phase, the C4.5 classifier is used to generate talent performance knowledge from yearly performance evaluation

database. The evaluation on other classifiers, such as neural network and nearest neighbor using the same training set, has been discussed in our previous work [24, 25]. In that study, the proposed classifier for the training dataset is C4.5 which is from the decision tree family. Finally, the generated classification rules from C4.5 classifier for talent performance have been analyzed to determine interesting attributes and the correctly classified data using unseen data.

This paper is organized in five sections. The first part is the introduction, followed by the second section which describes some related work on managing human talent, data mining in HR, classification for prediction, the decision tree techniques and C4.5 classifier for classification and prediction. The third section discusses the research method, while section 4 provides some results and discussion. Finally, section 5 ends with the concluding remarks and future research directions.

II. RELATED WORKS

A. Managing Human Talent

In any organization, talent management has become an increasingly crucial approach in HR functions. Talent is considered as the capability of any individual to make a significant difference to the current and future performance of the organization [26]. Managing an organization's talents involves human resource planning that emphasizes processes for managing people in organization. Besides that, talent management can be defined as a process to ensure leadership continuity in key positions and encourages individual advancement; and decision to manage supply, demand and flow of talent through human capital engine [27].

Talent management is very crucial and needs some attention from HR professionals. TP Track Research Report has found that 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 [19]. In this study, we attempt to focus on one of the talent management challenges i.e. to identify an existing talent regarding the key talent in an organization by predicting their performance. For that reason, we use the past data from the employee database to implement the classification and prediction process. Talent management process consists of recognizing the key talent areas in the organization, identifying the people in the organization who constitute key talent, and conducting development activities for the talent pool to retain and engage them and also have them ready to move into more significant roles [27] as illustrated in Fig.1. These processes involve HR activities that need to be integrated into an effective system such as decision support system [28].

B. Data Mining for Talent Management

Data mining is part of the process in Knowledge Discovery in Database (KDD). Data mining is a step in KDD and currently receives great attention and is recognized as a newly emerging analysis tool [20]. 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 amount of data and the important needs for turning such data into useful information and knowledge [29]. Data mining has several tasks such as classification and prediction; concept description; association; cluster analysis; outlier analysis; trend and evaluation analysis; statistical analysis and others. Computer application interfaces with data mining tool can help executives to make more informative and objective decisions. Besides that, it can help managers to retrieve, summarize and analyze related data to make wiser and more informative decisions. There are very few

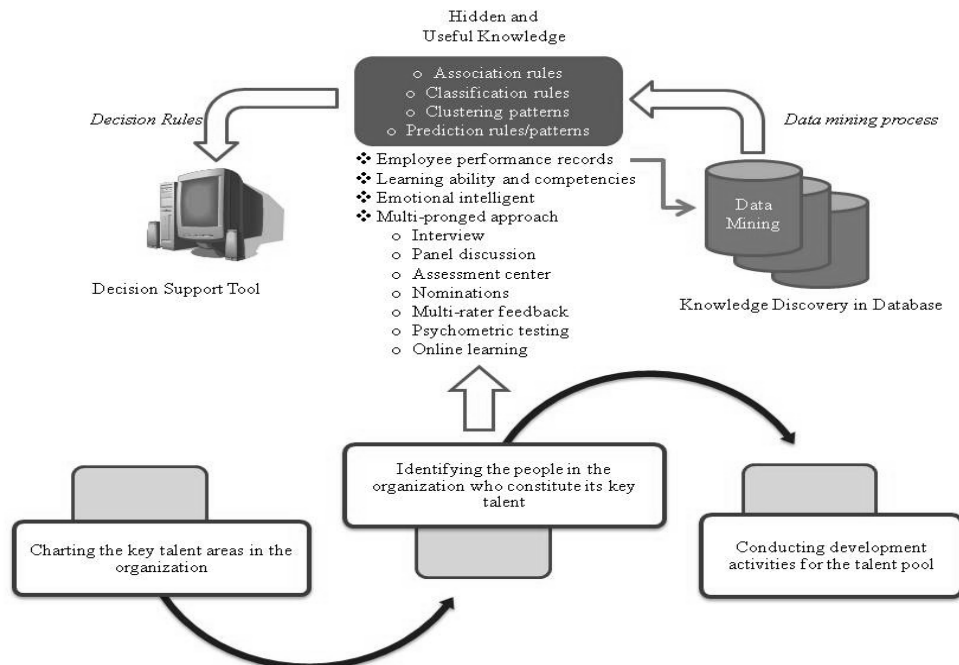


Figure 1. Human Talent Prediction using Data Mining Technique

studies related to prediction application in HR using this approach. However, this approach is quite popular in HR personnel selection problems. From the literature study, prediction applications in HRM are infrequent and there are some examples such as to predict the length of service, sales premiums, to achieve persistence indices of insurance agents and to analyze mis-operation behaviors of the operators [18]. Over the years, data mining has involved in various techniques including statistics, SVM, AIS, neural network, decision tree, genetic algorithm, and visualization techniques. Data mining has been applied in many fields such as finance, marketing, manufacturing, health care, customer relationship and etc. However, its application in HRM is rare [18].

Recently, with the new demand and increased visibility, HRM seeks a more strategic role by turning to data mining methods [17]. This can be done by identifying generated patterns from the existing data in HR databases as useful knowledge. Thus, this study concentrates on identifying the patterns that relate to the talent. The patterns can be generated by using some of the major data mining techniques, such as the clustering technique which is used to list the employees with similar characteristics, to group the performances and etc. From the association technique, patterns that are discovered can be used to associate the employee's profile for the most appropriate program/job, associated with employee's attitude to performance and etc. In classification and prediction, the pattern can be used to predict the percentage accuracy in employee's performance, behavior, and attitudes, predict the performance progress throughout the performance period, and also identify the best profile for different employee and etc. [30].

The matching of data mining problems and talent management needs is very crucial. Therefore, it is very important to determine the suitable data mining techniques.

In HRM, there are some interests on solving HRM problems using data mining approach [17, 31]. There are very few discussions on the uses of data mining related to employee's performance prediction, project assignment, employee's recruitment and many others. Due to these reasons, this study attempts to use the data mining approach for employee's performance prediction as one of the methods to predict the human talent in an organization. The purpose of this study is to determine the employees' performance by predicting their performance based on the past experience knowledge through previous performance evaluation data. In this study, the classification technique will be used for human talent prediction.

C. Classification for Prediction

Classification and prediction are among the methods that can produce intelligent decision. Currently, many classification and prediction methods have been proposed by researchers in machine learning, pattern recognition, and statistics. In this study, we are focusing on classification methods in data mining as part of machine learning process. Classification and prediction in data mining are two forms of data analysis that can be used to extract models to describe important data classes or to predict future data trends [32] as shown in Fig. 2.

The classification process has two phases; the first phase is learning process where the training data are analyzed by the classification algorithm. Learned model or classifier is represented in the form of classification rules. The second phase is classification process, where the test data are used to estimate the accuracy of classification model or classifier. If the accuracy is considered acceptable, the model can be applied to the new data to know the prediction result. There are many techniques that can be used for classification such as decision tree, Bayesian methods, Bayesian network, rule-based algorithms, neural network, support vector machine,

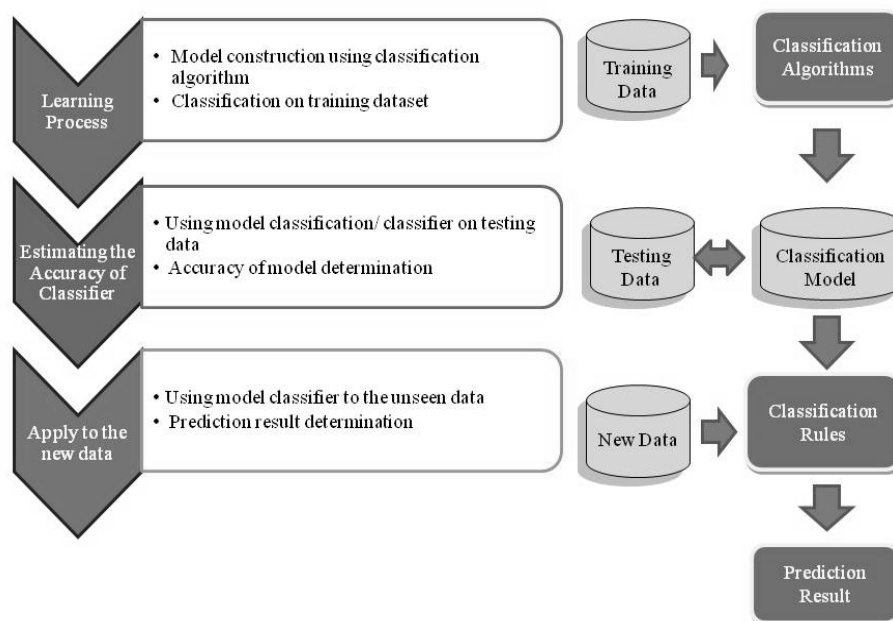


Figure 2. Classification Process in Data Mining

association rule mining, k-nearest-neighbor, case-based reasoning, genetic algorithms, rough sets and fuzzy logic. In this study, our discussion focuses on the three main classification techniques i.e. decision tree, neural network and k-nearest-neighbor. Decision tree and neural network are found useful in developing predictive models in many fields[20].

D. Decision Tree Techniques

Decision tree can produce a model with rules that are human-readable and interpretable. The classification task using decision tree technique can be performed without complicated computations and the technique can be used for both continuous and categorical variables. This technique is suitable for predicting categorical outcomes and less appropriate for application with time series data [20]. Decision tree classifiers are quite popular techniques because the construction of tree does not require any domain expert knowledge or parameter setting, and is appropriate for exploratory knowledge discovery. Currently, there are many research that employed decision tree techniques such as in electricity energy consumption [20], prediction of breast cancer [23], accident frequency [33], personnel selection [18], job attitudes [34] and others. It is stated that, the decision tree is among the powerful classification algorithms [20-23]. Some of decision tree classifiers are C4.5/C5.0/J4.8, NBTree, SimpleCart, REPTree, BFTree and others.

E. C4.5 Decision Tree Classifier

The C4.5 technique is one of the decision tree families that can produce both decision tree and rule-sets; and construct a tree for the purpose of improving prediction accuracy [21, 23]. Besides that, C4.5 models are easy to understand as the rules that are derived from the technique have a very straightforward interpretation. The C4.5 / C5.0 / J48 classifier is among the most popular and powerful decision tree classifiers [20-23]. C5.0 and J48 are the improved versions of C4.5 algorithms. WEKA toolkit package has its own version known as J48. J48 is an optimized implementation of C4.5 rev. 8. C4.5 creates an initial tree using the divide-and-conquer algorithm as described below [35]:

- If all the cases in S belong to the same class or S is small, the tree is a leaf labeled with the most frequent class in S.
- However, choose a test based on single attribute with two or more outcomes. Make this test the root of the tree with one branch for each outcome of the test, partition S into corresponding subsets S_1, S_2, \dots according to the outcome for each case, and apply to the same procedure recursively to each subset.

Usually there are many tests that can be selected in the last step. C4.5 uses two heuristic criteria to rank possible tests: information gain that uses attribute selection measure, which minimizes the total entropy of the subset $\{S_i\}$, and the default gain ratio that divides information gain by the information provided by the test outcomes. The information gain algorithm is described as the function Gain (A), which is shown below:

- Select the attribute with the highest information gain
- S contains s_i tuples of class C_i for $i = \{1, \dots, m\}$

- Information measure or expected information is required to classify any arbitrary tuple:

$$I(s_1, s_2, \dots, s_m) = - \sum_{i=1}^m \frac{s_i}{s} \log_2 \frac{s_i}{s} \tag{1}$$

- Entropy of attribute A with values $\{a_1, a_2, \dots, a_v\}$:

$$E(A) = \sum_{j=1}^v \frac{s_{1j} + \dots + s_{mj}}{s} I(s_{1j}, \dots, s_{mj}) \tag{2}$$

- Information gain means how much can be gained by branching on attribute A:

$$\text{Gain}(A) = I(s_1, s_2, \dots, s_m) - E(A) \tag{3}$$

III. METHODOLOGY

In this study, we attempt to discover employees' performance patterns from the existing employees' performance data using the decision tree classification techniques. The techniques are chosen based on the common techniques for classification and prediction in data mining. The decision tree is a 'divide-and-conquer' approach from a set of independent instances. In this experiment, we used C4.5 classifier which comes from the decision tree family and has been evaluated in our previous work [25, 36]. The input variables for the process are performance factors for selected attributes; and the outcome is the employees' performance represented by the status of recommendation for promotion, whether "yes" or "no" as shown in Table I.

The attributes for training dataset are selected based on the related factors for employee performance, as illustrated in Fig.3. These attributes are extracted from the individual factors component i.e. work outcome; knowledge and skill; individual quality; and activities and contribution. In this study, the performance factors are based on the performance appraisal standard used by the Malaysian public sector organizations. Besides the performance factors, some background information is also considered as part of the attributes for the dataset.

TABLE I. ATTRIBUTE DESCRIPTION

Attribute/ Factor	Description
Category	P – Professional, S - Support Staff
Gender	Male and Female
Qualification	Doctorate, Master, Bachelor, Diploma and Certificate
PK _{1...6}	Work Outcome (50%)
PM _{1...6}	Knowledge and Skill (25%)
KP _{1...6}	Individual Quality (20%)
KS _{1...6}	Activities and Contribution (5%)
YEAR _{1...6}	Evaluation mark (100%)
Target/Class	Recommendation for promotion (Yes or No)

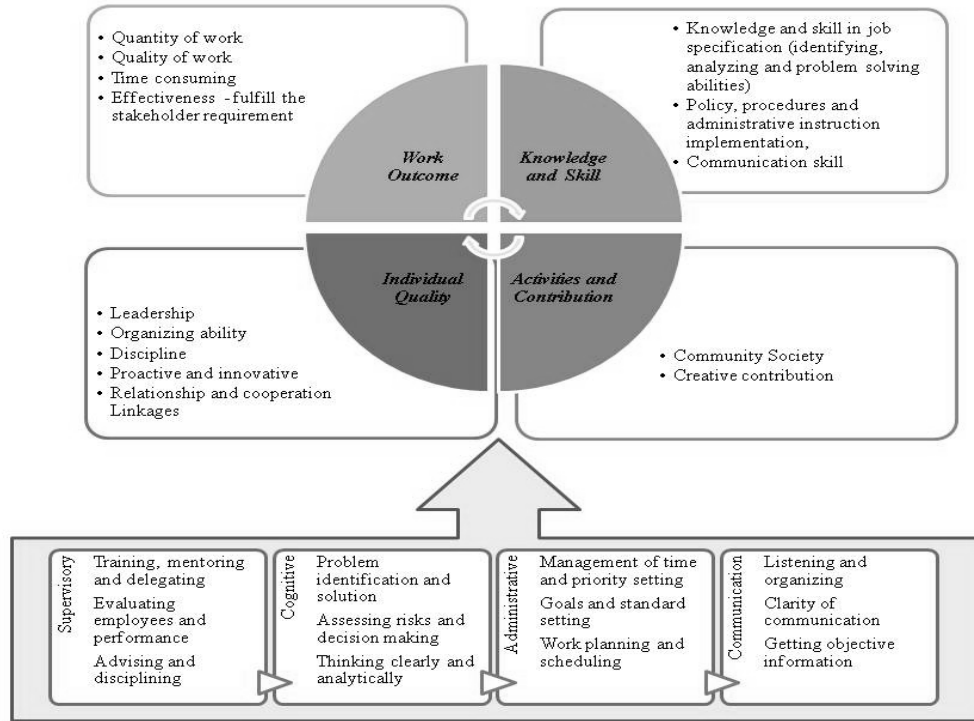


Figure 3. Human Talent Prediction using Data Mining Techniques

The aim of this study is to generate the forecasting model that contains classification rules for human talent prediction. The classification rules also show us about the interesting or important attributes for the dataset. The forecasting model will be used to determine whether the employee is recommended for promotion or not based on his/her performance. In this experiment, the training dataset contained 33 related attributes from background information and performance factors which are demonstrated in Table I. The initial dataset contains 655 records of performance evaluation marks from the six years (2003-2008). Each record holds evaluation marks for selected factors and the total mark for each of the years. The data mining tools used are WEKA and ROSETTA toolkit. The flow of the research process is shown in Fig.4.

This study has three phases; the first phase is the data collection process which involved the data cleaning and data preprocessing. The second phase is to generate the classification rules using C4.5 classifier for the training dataset. In this case, we use all the selected attributes defined in Table I. The C4.5 classifier produced the analysis of the training dataset and the classification rules. In the preprocessing phase of the experiment, the initial dataset is divided into two sets of data. The first data set is the training dataset which contains 590 data and the second one is testing dataset or unseen data. The testing dataset comprises of 10% records extracted from the initial data. The third phase of experiment is the evaluation and interpretation of the classification rules using the unseen data. The purpose of the evaluation process is to determine the accuracy of classification rules for prediction and to identify the important or interesting attributes and rules. The hidden knowledge from the performance evaluation is embedded into the decision support system for employees' performance prediction. Thus, the intelligent decision

support system can be used to predict whether the employee is recommended for promotion or not.

Figure 4. Research Phases

IV. RESULTS AND DISCUSSIONS

The classification rules generated from C4.5 classifier is human readable and easy to understand which do not require any domain expert knowledge or parameter setting. This technique can produce both the decision tree and rule-sets and can construct a tree for the purpose of improving the prediction accuracy [21, 23]. On top of that, C4.5 model produces easy to understand rules with a very straightforward interpretation. As stated in [1-4], the C4.5/C5.0/J4.8 classifier is among the popular and powerful decision tree classifiers. The extracted summary of the analysis for C4.5 classifier on the performance evaluation training dataset from WEKA toolkit is shown below:

<i>Correctly Classified Instances</i>	561	95.0847 %
<i>Incorrectly Classified Instances</i>	29	4.9153 %
<i>Kappa statistic</i>	0.8489	
<i>Mean absolute error</i>	0.0861	
<i>Root mean squared error</i>	0.2075	
<i>Relative absolute error</i>	25.3139 %	
<i>Root relative squared error</i>	50.3528 %	
<i>Total Number of Instances</i>	590	

In addition, Fig. 5 shows part of the decision tree classification model using C4.5 classifier. In this case, the leaves in the tree represent the classification result or the target class. As mentioned earlier, the classification's target is the recommendation for promotion. The C4.5 classifier uses two heuristic criteria to rank possible tests: information gained by using attribute selection measure and the default gained ratio that divides information gain by the

information provided by the test outcomes. With these criteria, C4.5 classifier can also be used to determine the important or interesting attributes from the dataset. In this study, the important attributes are identified through the number of hits for each of the attributes from generated classification rules. The number of hits for all selected attributes is shown in Table II.

TABLE II. INTERESTING ATTRIBUTES

Attributes	Hits	Attributes	Hits
Category	1	Contribution3	2
Gender	2	Outcome4	2
Qualification	1	Knowledge4	2
Outcome1	2	Personal4	1
Knowledge1	2	Contribution4	1
Personal1	2	Knowledge5	1
Contribution1	2	Contribution5	1
YEAR1	2	Outcome6	2
Outcome2	2	Knowledge6	2
Knowledge2	4	Personal6	1
Contribution2	1	Contribution6	2
Knowledge3	1		

In this experiment, from 33 selected attributes, only 23 attributes are hit in the classification rules. As shown in

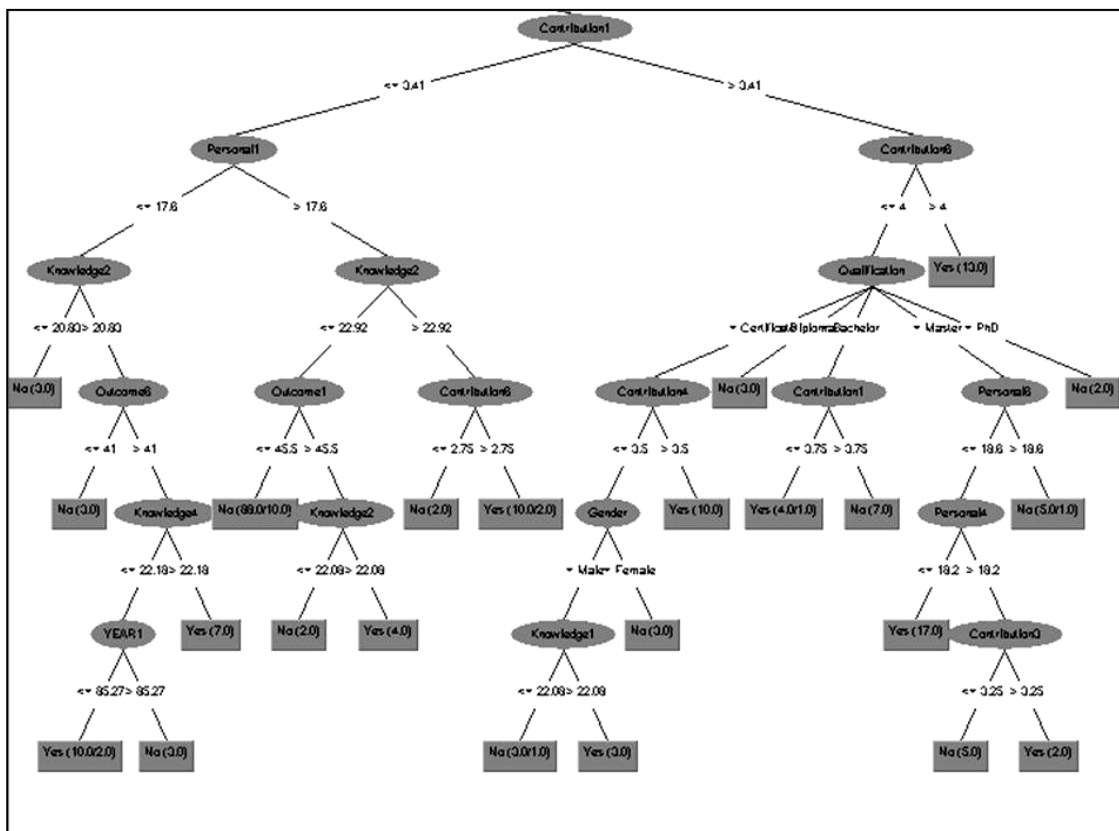


Figure 5. Sample of Rules using C4.5 Classifier

Table II, the result reveals that the *Knowledge2* attribute has the highest hits as compared to other attributes. In this case, this result indicates the importance of the attribute for the training dataset. The *knowledge2* attribute represents the factor of knowledge and skill for year 2 performance evaluation marks. This experiment result also indicates the important of knowledge and skill factor for the employees' performance, where the knowledge and skill attributes are all listed in the importance of attributes in Table II. Another important analysis from the experiment is the analysis of the generated rules. The C4.5 classifier generated 43 classification rules from 590 training datasets. The sample decision tree is illustrated in Figure 3 and the sample rules that are extracted from the tree are shown in Table III.

TABLE III. SAMPLE OF CLASSIFICATION RULES

Rule No	Classification Rules	Result
1	IF (<i>Knowledge5</i> <= 22.26) AND IF (<i>Outcome4</i> <= 45) AND IF (<i>Contribution2</i> <= 3.75) AND IF (<i>Contribution1</i> <= 3) AND IF (<i>Knowledge6</i> <= 22.05) AND IF (<i>Contribution3</i> <= 3.25) AND IF (<i>Gender</i> = Male) AND IF (<i>Outcome2</i> <= 43.5)	NO
2	IF (<i>Knowledge5</i> <= 22.26) AND IF (<i>Outcome4</i> <= 45) AND IF (<i>Contribution2</i> <= 3.75) AND IF (<i>Contribution1</i> <= 3) AND IF (<i>Knowledge6</i> <= 22.05) AND IF (<i>Contribution3</i> <= 3.25) AND IF (<i>Gender</i> = Male) AND IF (<i>Outcome2</i> > 43.5)	YES
3	IF (<i>Knowledge5</i> <= 22.26) AND IF (<i>Outcome4</i> <= 45) AND IF (<i>Contribution2</i> <= 3.75) AND IF (<i>Contribution1</i> <= 3) AND IF (<i>Knowledge6</i> <= 22.05) AND IF (<i>Contribution3</i> <= 3.25) AND IF (<i>Gender</i> = Female)	NO

Table IV shows the analysis of rules by looking at the number of hits for each rule applied on the unseen or testing data (65 data). In this case, rule no 8 has the highest number of hits as compared to the other rules. From that, we can conclude that the rule is important or interesting. However, further experiments are needed to verify this finding. Although the finding seems very interesting, further studies are needed to validate and explore more on the result. The experiment related to the relevancy analysis should be conducted to determine the effects on the accuracy of the classifier. Besides that, the experiment for attribute reduction should also be performed as an approach to reduce the processing time and space while improving the accuracy of the classification model.

In this study, we observed the great potential of C4.5 as a classification technique for employee's performance prediction. In the next stage of data mining classification technique analysis, C4.5 classifier should be further analyzed with other datasets to justify the suitability of the C4.5 classifier for the performance evaluation dataset. Furthermore, the other decision tree classification algorithms should also be tested in order to validate these findings. For that reason, in the next experiment, other decision tree

techniques such as NBTree, SimpleCart, REPTree, BFTree and many others should be tested to support the technique.

TABLE IV. RULE ANALYSIS

Rule	Hits	Rule	Hits	Rule	Hits	Rule	Hits
1		12		23		34	
2	1	13		24		35	3
3	1	14	4	25	2	36	1
4	4	15	1	26	2	37	1
5	1	16		27		38	3
6		17		28	1	39	1
7		18		29		40	1
8	22	19	1	30	1	41	4
9		20	1	31		42	1
10		21		32		43	5
11	3	22		33			

Table V shows the rules evaluation where 52 records matched the rules which is 77% from the total number of records in the new data set. This result indicates the accuracy of the classification for the new dataset.

TABLE V. RULE EVALUATION

Status	Number of Data	Accuracy(%)
Correctly Classified	52	77
Incorrectly Classified	13	23

V. CONCLUSION

This article has described the significance of the study on the use of data mining classification techniques for employees' performance prediction. However, there should be more data mining classification techniques applied to the different problem domains in HR field of research to broaden the horizon of academic and practice work on data mining in HR. Besides that, other data mining techniques such as SVM, Fuzzy logic and AIS should also be considered for future work on classification techniques using the same dataset.

In this study, as we can see from result analysis, C4.5 classifier has a great potential for performance prediction. The generated classification rules can be used to predict the performance of an employee whether he/she has potential to be promoted or not, based on his/her performance.. Currently, this research is at the stage of system development where the classification rules need to be embedded into the decision support system which is known as Intelligent Decision Support System (IDSS). The IDSS is a substantial of intelligent system which can help the decision makers to determine their potential employee for promotion or may be for other tasks. This intelligent system

will be discussed in our next study. In conclusion, the ability to continuously change and obtain new understanding of the classification and prediction in HR research has become the major contribution to data mining in HRM.

ACKNOWLEDGMENT

The authors are indebted to Madam Hanita Yusuf, Assistant Registrar at Universiti Teknologi MARA Terengganu for her permission to use the existing employees' evaluation marks as training data set in this study.

REFERENCES

- [1] Y. Feng, et al., "Knowledge discovery in traditional Chinese medicine: State of the art and perspectives," *Artificial Intelligence in Medicine*, vol. 38, pp. 219-236, 2006.
- [2] E. Frank, Hall, M., et al., "Data mining in bioinformatics using Weka," *Bioinformatics Application Note*, vol. 20, pp. 2479-2481, 2004.
- [3] C. Combes, Meskens, N., Rivat., C. & Vandamme J.P., "Using a KDD process to forecast the duration of surgery," *International Journal of Production Economics*, vol. 112, pp. 279-293, 2008.
- [4] C. L. Chang, & Chen, C.H., "Applying decision tree and neural network to increase quality of dermatologic diagnosis," *Expert Systems with Applications*, vol. 36, pp. 4035-4041, 2009.
- [5] A. S. Chang, & Leu, S.S., "Data mining model for identifying project profitability variables," *International Journal of Project Management*, vol. 24, pp. 199-206, 2006.
- [6] B. Ekasingh, Ngamsomsuke, K., et al., "A data mining approach to simulating farmers' crop choices for integrated water resources management," *Journal of Environmental Management*, vol. 77, pp. 315-325, 2005.
- [7] R. Arbel, & Rokach, L., "Classifier evaluation under limited resources," *Pattern Recognition Letters*, vol. 27, pp. 1619-1631, 2006.
- [8] F. E. Ciarapica, & Giacchetta. G., "Classification and prediction of occupational injury risk using soft computing techniques: An Italian study," *Safety Science*, vol. 47, pp. 36-49, 2009.
- [9] V. Cho, & Ngai, E.W.T., "Data mining for selection of insurance sales agents," vol. 20, pp. 123-132, 2003.
- [10] S. H. Liao, Chen, Y.N., & Tseng, Y.Y., "Mining demand chain knowledge of life insurance market for new product development," *Expert Systems with Applications*, vol. 36, pp. 9422-9437, 2009.
- [11] S. T. Li and L. Y. Shue, "Data mining to aid policy making in air pollution management," *Expert Systems with Applications*, vol. 27, pp. 331-340, 2004.
- [12] W. S. D. Chen, Y.K., "Using neural networks and data mining techniques for the financial distress prediction model " *Expert System with Applications*, vol. 36, pp. 4075-4086, 2009.
- [13] I. Bose, & Mahapatra, R.K., "Business data mining - a machine learning perspective " *Information & Management*, vol. 39, pp. 211-225, 2001.
- [14] C. Rygielski, J. C. Wang, and D. C. Yeh, "Data mining techniques for customer relationship management," *Technology in Society*, vol. 24, pp. 483-502, 2002.
- [15] C. Romero and S. Ventura, "Educational data mining : A survey from 1995 to 2005," *Expert Systems with Applications*, vol. 33, pp. 135-146, 2007.
- [16] J. Ranjan and K. Malik, "Effective educational process: A data mining approach," *VINE: The Journal of Information and Knowledge Management Systems*, vol. 37, pp. 502-515, 2007.
- [17] J. Ranjan, "Data Mining Techniques for better decisions in Human Resource Management Systems," *International Journal of Business Information Systems*, vol. 3, pp. 464-481, 2008.
- [18] C. F. Chien 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*, vol. 34, pp. 380-290, 2008.
- [19] A TP Track Research Report "Talent Management: A State of the Art," Tower Perrin HR Services 2005.
- [20] G. K. F. Tso and K. K. W. Yau, "Predicting electricity energy consumption: A comparison of regression analysis, decision tree and neural networks," *Energy*, vol. 32, pp. 1761-1768, 2007.
- [21] I. Becerra-Fernandez, S. H. Zanakis, and S. Walczak, "Knowledge discovery techniques for predicting country investment risk," *Computers & Industrial Engineering*, vol. 43, pp. 787-800, 2002.
- [22] P. R. Kumar and V. Ravi, "Bankruptcy prediction in banks and firms via statistical and intelligent techniques : A review," *European Journal of Operational Research*, vol. 180, pp. 1-28, 2007.
- [23] D. Delen, G. Walker, and A. Kadam, "Predicting breast cancer survivability: A comparison of three data mining methods," *Artificial Intelligent in Medicine*, vol. 34, pp. 113-127, 2005.
- [24] J. Hamidah, H. Abdul Razak, and O. Zulaiha Ali, "Classification Techniques for Talent Forecasting in Human Resource Management " in *Advanced Data Mining and Application*, R. H. Q. Yang, J. P. J. Gama, and X. Meng, Eds. Beijing, China: Springer-Verlag Berlin Heidelberg, 2009, pp. 496-503.
- [25] J. Hamidah, H. Abdul Razak, and A. O. Zulaiha, "Classification for Talent Management using Decision Tree Induction Techniques," in *2nd Data Mining and Optimization Seminar (DMO'09)*, Bangi, Selangor, 2009, pp. 15-20.
- [26] M. Lynne, "Talent Management Value Imperatives : Strategies for Execution," The Conference Board 2005.
- [27] I. Cubbingham, "Talent Management : Making it real," *Development and Learning in Organizations*, vol. 21, pp. 4-6, 2007.
- [28] CHINA UPDATE, "HR News for Your Organization : The Tower Perrin Asia Talent Management Study," 2007.
- [29] J. Han and M. Kamber, *Data Mining : Concepts and Techniques*. San Francisco: Morgan Kaufmann Publisher, 2006.
- [30] H. Jantan, A. R. Hamdan, and Z. A. Othman, "Data Mining Techniques for Performance Prediction in Human Resource Application," in *1st Seminar on Data Mining and Optimization*, Selangor, 2008, pp. 41-49.
- [31] H. Jantan, A. R. Hamdan, and Z. A. Othman, "Classification Techniques for Talent Forecasting in Human Resource Management " in *5th International Conference on Advanced Data Mining and Application (ADMA)*, Beijing, China, 2009, pp. 496-503.
- [32] J. Han and M. Kamber, *Data Mining: Concepts and Techniques*. San Francisco: Morgan Kaufmann Publisher, 2006.
- [33] L. Y. Chang and W. C. Chen, "Data mining of tree-based models to analyze freeway accident frequency," *Journal of Safety Research* vol. 36, pp. 365-375, 2005.
- [34] K. Y. Tung, I. C. Huang, S. L. Chen, and C. T. Shih, "Mining the Generation Xer's job attitudes by artificial neural network and decision tree - empirical evidence in Taiwan," *Expert Systems and Applications*, vol. 29, pp. 783-794, 2005.
- [35] X. Wu, V. Kumar, J. R. Quinlan, J. Ghosh, Q. Yang, H. Motoda, G. J. McLachlan, A. Ng, B. Liu, P. S. Yu, Z.-H. Zhou, M. Steinbach, D. J. Hand, and D. Steinberg, "Top 10

algorithms in data mining," *Knowledge Information System*, 2008.

- [36] H. Jantan, A. R. Hamdan, Z. A. Othman, and M. Puteh, "Applying Data Mining Classification Techniques for Employee's Performance Prediction," in *Knowledge Management 5th International Conference (KMICe2010)*, Kuala Terengganu, Terengganu Malaysia, 2010, pp. 645-652.

AUTHORS PROFILE



Hamidah Jantan received the first degree in Computer Science in 1989 from Universiti Teknologi Malaysia (UTM), Kuala Lumpur, Malaysia. She obtained her master degree in Information Technology (Science and System Management) from Universiti Kebangsaan Malaysia (UKM) in 2002.

She is currently working towards her Ph.D. degree in the Faculty of Information Science and Technology (FTSM), Universiti Kebangsaan Malaysia (UKM), Malaysia. In addition, she is a senior lecturer at Universiti Teknologi MARA (UiTM) Terengganu. Her research interests include Intelligent System, Decision Support System and Data Mining technology.



Abdul Razak Hamdan is a Professor at Faculty of Information Science and Technology (FTSM), Universiti Kebangsaan Malaysia (UKM). He received his first degree in Science from Universiti Kebangsaan Malaysia (UKM) in 1975, his master degree in Computing Science from University of Newcastle Upon Tyne, England in 1977, and PhD degree in Artificial Intelligent from Loughborough University of Technology, England in 1987. His research interests include Artificial Intelligent, Decision Support System, Strategic Planning and Data mining.



Zulaiha Ali Othman is an associate Professor at Faculty of Information Science and Technology (FTSM), Universiti Kebangsaan Malaysia (UKM). She received her first degree in computer science from Universiti Kebangsaan Malaysia (UKM) in 1990, her master degree in Software Technology from University of Sheffield, UK in 1997, and PhD degree in Computing (Agent Oriented Methodology) from Sheffield Hallam University, UK, in 2003. Her research interests include Artificial Intelligent, Agent Technology and Data mining.