

Using Data Mining Techniques to Build a Classification Model for Predicting Employees Performance

Qasem A. Al-Radaideh

Department of Computer Information Systems,
Faculty of Information Technology and Computer Science
Yarmouk University, Irbid 21163, Jordan.

Eman Al Nagi

Department of Computer Science, Faculty of Information
Technology, Philadelphia University, Amman, Jordan.

Abstract— Human capital is of a high concern for companies' management where their most interest is in hiring the highly qualified personnel which are expected to perform highly as well. Recently, there has been a growing interest in the data mining area, where the objective is the discovery of knowledge that is correct and of high benefit for users. In this paper, data mining techniques were utilized to build a classification model to predict the performance of employees. To build the classification model the CRISP-DM data mining methodology was adopted. Decision tree was the main data mining tool used to build the classification model, where several classification rules were generated. To validate the generated model, several experiments were conducted using real data collected from several companies. The model is intended to be used for predicting new applicants' performance.

Keywords- Data Mining; Classification; Decision Tree; Job Performance.

I. INTRODUCTION

Human resource has become one of the main concerns of managers in almost all types of businesses which include private companies, educational institutions and governmental organizations. Business Organizations are really interested to settle plans for correctly selecting proper employees. After hiring employees, managements become concerned about the performance of these employees were management build evaluation systems in an attempt to preserve the good-performers of employees (Chein and Chen, 2006).

Data mining is a young and promising field of information and knowledge discovery (Han et al., 2011). It started to be an interest target for information industry, because of the existence of huge data containing large amounts of hidden knowledge. With data mining techniques, such knowledge can be extracted and accessed transforming the databases tasks from storing and retrieval to learning and extracting knowledge.

Data mining consists of a set of techniques that can be used to extract relevant and interesting knowledge from data. Data mining has several tasks such as association rule mining, classification and prediction, and clustering. Classification techniques 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 models that are used to predict future data trends. There are several algorithms for data classification such as decision tree and Naïve Bayes classifiers. With classification, the generated model will be able to predict a class for given data depending on previously learned information from historical data.

Decision tree 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 (Han et al., 2011).

Naïve Bayes classifier is another classification technique that is used to predict a target class. It depends in its calculations on probabilities, namely Bayesian theorem. Because of this use, results from this classifier are more accurate and effective, and more sensitive to new data added to the dataset (Han et al., 2011).

Several studies used data mining for extracting rules and predicting certain behaviors in several areas of science, information technology, human resources, education, biology and medicine.

For example, Beikzadeh and Delavari (2004) used data mining techniques for suggesting enhancements on higher educational systems. Al-Radaideh et al. (2006) also used data mining techniques to predict university students' performance. Many medical researchers, on the other hand, used data mining techniques for clinical extraction units using the enormous patients data files and histories, Lavrac (1999) was one of such researchers. Mullins et al. (2006) also worked on patients' data to extract disease association rules using unsupervised methods.

Karatepe et al. (2006) defined the performance of a frontline employee, as his/her productivity comparing with his/her peers. Schwab (1991), on the other hand, described the performance of university teachers included in his study, as the number of researches cited or published. In general,

performance is usually measured by the units produced by the employee in his/her job within the given period of time.

Researchers like Chein and Chen (2006) have worked on the improvement of employee selection, by building a model, using data mining techniques, to predict the performance of newly applicants. Depending on attributes selected from their CVs, job applications and interviews. Their performance could be predicted to be a base for decision makers to take their decisions about either employing these applicants or not.

Previous studies specified several attributes affecting the employee performance. Some of these attributes are personal characteristics, others are educational and finally professional attributes were also considered. Chein and Chen (2006) used several attributes to predict the employee performance. They specified age, gender, marital status, experience, education, major subjects and school tires as potential factors that might affect the performance. Then they excluded age, gender and marital status, so that no discrimination would exist in the process of personal selection. As a result for their study, they found that employee performance is highly affected by education degree, the school tire, and the job experience.

Kahya (2007) also searched on certain factors that affect the job performance. The researcher reviewed previous studies, describing the effect of experience, salary, education, working conditions and job satisfaction on the performance. As a result of the research, it has been found that several factors affected the employee's performance. The position or grade of the employee in the company was of high positive effect on his/her performance. Working conditions and environment, on the other hand, had shown both positive and negative relationship on performance. Highly educated and qualified employees showed dissatisfaction of bad working conditions and thus affected their performance negatively. Employees of low qualifications, on the other hand, showed high performance in spite of the bad conditions. In addition, experience showed positive relationship in most cases, while education did not yield clear relationship with the performance.

In their study, Salleh et al. (2011) have tested the influence of motivation on job performance for state government employees in Malaysia. The study showed a positive relationship between affiliation motivation and job performance. As people with higher affiliation motivation and strong interpersonal relationships with colleagues and managers tend to perform much better in their jobs.

Jantan et al. (2010) had discussed in their paper Human Recourses (HR) system architecture to forecast an applicant's talent based on information filled in the HR application and past experience, using Data Mining techniques. The goal of the paper was to find a way to talent prediction in Malaysian higher institutions. So, they have specified certain factors to be considered as attributes of their system, such as, professional qualification, training and social obligation. Then, several data mining techniques (hybrid) where applied to find the prediction rules. ANN, Decision Tree and Rough Set Theory are examples of the selected techniques.

The same authors, Jantan et al. (2010b) have used decision tree C4.5 classification algorithm to predict human talent in

HRM, by generating classification rules for the historical HR records, and testing them on unseen data to calculate accuracy. They intend to use these rules in creating a DSS system that can be used by managements to predict employees' performance and potential promotions.

Generally, this paper is a preliminary attempt to use data mining concepts, particularly classification, to help supporting the human resources directors and decision makers by evaluating employees' data to study the main attributes that may affect the employees' performance. The paper applied the data mining concepts to develop a model for supporting the prediction of the employees' performance. In section 2, a complete description of the study is presented, specifying the methodology, the results, discussion of the results.

II. BUILDING THE CLASSIFICATION MODEL

The main objective of the proposed methodology is to build the classification model that tests certain attributes that may affect job performance. To accomplish this, the CRISP-DM methodology (Cross Industry Standard Process for Data Mining) (CRISP-DM, 2007) was used to build a classification model. It consists of five steps which include: Business understanding, data understanding, data preparation, modeling, evaluation and deployment.

A. Data Classification Preliminaries

In general, data classification is a two-step process. In the first step, which is called the learning step, a model that describes a predetermined set of classes or concepts is built by analyzing a set of training database instances. Each instance is assumed to belong to a predefined class. In the second step, the model is tested using a different data set that is used to estimate the classification accuracy of the model. If the accuracy of the model is considered acceptable, the model can be used to classify future data instances for which the class label is not known. At the end, the model acts as a classifier in the decision making process. There are several techniques that can be used for classification such as decision tree, Bayesian methods, rule based algorithms, and Neural Networks.

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. Decision tree can produce a model with rules that are human-readable and interpretable. Decision Tree has the advantages of easy interpretation and understanding for decision makers to compare with their domain knowledge for validation and justify their decision. Some of decision tree classifiers are C4.5/C5.0/J4.8, NBTree, and others.

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. The C4.5 / C5.0 / J48 classifier is among the most popular and powerful decision tree classifiers. C4.5 creates an initial tree using the divide-and-conquer algorithm. The full description of the algorithm can be found in any data mining or machine learning books such as (Han et al., 2011) and (Witten et al., 2011).

WEKA toolkit (Witten et al., 2011) is a widely used toolkit for machine learning and data mining originally developed at the University of Waikato in New Zealand. It contains a large collection of state-of-the-art machine learning and data mining algorithms written in Java. WEKA contains tools for regression, classification, clustering, association rules, visualization, and data pre-processing. WEKA has become very popular with academic and industrial researchers, and is also widely used for teaching purposes. WEKA toolkit package has its own version known as J48. J48 is an optimized implementation of C4.5 rev. 8.

B. Data Collection Process and Data Understanding

When the idea of the study came in to mind, it was intended to apply a classification model for predicting performance depending on a dataset from a certain IT company. So that any other factors regarding the working environment, conditions, management and colleagues would have similar effect on all employees, and so the effect of collected attributes would be more apparent and easier to classify. Unfortunately, data collected from the first IT Company was not enough to be the base of such a classification model. In this case, another attempt was taken to collect another group of data from another IT company. In order to collect the required data, a questionnaire was prepared and distributed either by email or manually to the employees of both companies. Then, it was further distributed on the internet, to be filled by employees working in any IT company. The questionnaire was filled by 130 employees, 37 from the first IT Company, 38 from the second one, and the rest from several other companies using the internet questionnaire.

Several attributes have been asked for in the questionnaire that might predict the performance class. The list of the collected attributes is presented in Table 1.

C. Data Preparation

After the questionnaires were collected, the process of preparing the data was accomplished. First, the information in the questionnaires has been transferred to Excel sheets. Then, the types of data has been reviewed and modified. Some attributes like experience years and service period, have been entered in continuous values. So, they were modified to be illustrated by ranges. Other attributes like specialization, job title and rank, have been generalized to include fewer discrete values than they already have. For example, in specialization, there were values like electrical engineering and computer engineering, they have been considered as one value, engineering. MIS and CIS were considered as IT, and so on.

These files are prepared and converted to (arff) format to be compatible with the WEKA data mining toolkit (Witten et al., 2011), which is used in building the model.

As mentioned previously, the data has been divided into three datasets. The first one includes the data of the first IT company employees. The second includes the data of the second IT company employees. The third one includes all data collected from the three sources. Each dataset has two arff files containing its data, with the class attribute (performance). Each of these datasets was used in a separate experiment.

III. MODELING AND EXPERIMENTS

After the data has been prepared, the classification models have been built. Using the decision tree technique, a tree has been built for each of these experiments. In this technique, the gain ratio measure is used to indicate the weight of effectiveness of each attribute on the tested class, and accordingly the ordering of tree nodes is specified. The results are discussed in the following sections.

Referring to the discussion of earlier studies, and as described in Table (1), a group of attributes has been selected to be tested against their effectiveness on the employee performance.

These attributes consist of (1) Personal information such as: age, gender, marital status and number of kids (if any), (2) Education information such as: university type, general specialization, degree and grade, (3) Professional information such as: number of experience years, number of previous companies worked for, job title, rank, service period in the current company, salary, finding the working conditions uncomfortable and dissatisfaction of salary or rank. These attributes were used to predict the employee performance to be *accomplished, exceed or far-exceed*.

A. First Experiment (E1): Using the whole dataset (130 instances)

Three classification techniques have been applied on the dataset on hand to build the classification model. The techniques are: The decision tree with two versions, ID3 and C4.5 (J4.8 in WEKA), and Naïve Bayes classifier. For each experiment, accuracy was evaluated using 10-folds cross-validation, and hold-out method. Table 2 displays the accuracy percentages for each of these techniques.

The tree generated by ID3 algorithm was very deep, since it started by attribute JobTitle, which has 20 values.

The JobTitle has the maximum gain ratio, which made it the starting node and most effective attribute. Other attributes participated in the decision tree were UnivType, SalRange, ExpYears, Grade, Age, MStatue, Gender, GSpecial and Rank. Other attributes such as: PrevCo, Nkids, uncomworkcond, dissatsalrank and degree appeared in other parts of the decision tree.

TABLE I. FULL DESCRIPTION OF ATTRIBUTES USED FOR PREDICTING THE PERFORMANCE CLASS

Attribute	Description	Possible Values
*Age	Employee's Age	a, b, c, d, e
Gender	Employee's Gender	Male, Female
MStatue	Employee's Marital Status	Single, Married with kids, Married without kids, Other
NKids	No. of Kids	0, 1, 2, 3, 4

UnivType	The type of university of graduation	Public, Private
GSpecial	General Specialization	Business, IT, English Literature, Engineering, CS, Other
Degree	Employee Education Degree	Diploma, Bachelor, High Diploma, Master, PhD
Grade	Employee Graduation Grade	Excellent, Very good, Good, Acceptable, Other
Country	Country of University	This attribute was eliminated, since majority of the values were Jordan
**Expyears	No. of Working Experience Years	a, b, c, d, e
PrevCo	No. of Previous Companies the employee worked for	0,1,2,3,4,5
JobTitle	Employee's Job Title in the current company	Developer, Officer, QA, Data Entry, System Administrator, Office Manager, Technical Writer, Technical Manager, Software Engineering, Accountant, Infrastructure Engineer, Department Manager, software Architect, Analyst, Designer, Trainer, PM, Consultant, Customer Support, GM.
Rank	Employee's Rank in the current company	Junior, Senior, Team Leader, Manager, Architect
***ServPeriod	Service Period in the current company (in years)	a, b, c, d
Working hours	No of working hours	Full Time, Part Time, Other. This attribute was eliminated, since most instances has Full time value.
****SalRange	Employee's Range of Salary	a, b, c, d, e
UncomWorkcond	Working in uncomfortable conditions (in employee's perspective)	Yes, No
Dissatsalrank	Existence of dissatisfaction in either salary or rank	Yes, No
Performance	Employee's performance, either as informed or predicted. This is a class	Accomplish, Exceed, Far Exceed

Notes: * Age range: a) Under 25 b) 25-29 c) 30-34 d) 35-40 e) Over 40
 ** Experience Years Range: a) None b) Less than a year c) 1-5 d) More than 5 – 10 e) More than 10
 *** Service Period Range: a) Less than 1 year b) 1-5 c) More than 5-10 d) More than 10
 **** Salary Range: a) Less than 250 b) 250-499 c) 500-999 d) 1000-2000 e) Over 2000

TABLE II. ACCURACY PERCENTAGES FOR PREDICTING PERFORMANCE IN E1

Method	10-Fold Cross Validation	Hold-out (60%)
ID3	36.9%	36.5%
C4.5 (J4.8)	42.3%	48.1%
Naïve Bayes	40.7%	44.2%

The tree indicated that all these attributes have some sort of effect on the employee performance, but the most affective attributes were: JobTitle, UnivType and Age. Other hints could be extracted from the tree indicates that young employees have better performance than older ones. Wherever Gender is taken into consideration, Male employees have higher performance than Female. Moreover, employees with higher graduation grades have higher performance. Finally, employees with higher ranks have less performance giving indication that managers work less than less ranked employees.

The tree generated using the C4.5 algorithm also indicated that the JobTitle attribute is the most affective attribute. The Naïve Bayes classifier does not show the weights of each attribute included in the classification, but it has been used to be compared with the results generated from ID3 and C4.5 as was shown previously in Table 2. It can be noticed that the accuracy percentage ranges from approximately 36% to 45%, which are low percentages.

B. Second Experiment (E2): Using the dataset gathered from the first IT company (37 instances)

By using the same approach as in E1, Table 3 shows the prediction accuracy for each algorithm applied to this dataset.

TABLE III. ACCURACY PERCENTAGES FOR PREDICTING PERFORMANCE IN E2

Algorithm	10-Fold Cross Validation	Hold-out (60%)
ID3	37.8 %	26. 7%
C4.5 (J4.8)	48.6%	53.3%
Naïve Bayes	37.8%	46. 7%

Decision tree built by ID3 algorithm, showed different trend than the one generated for E1, since in this tree, the starting node was PrevCo, and the attributes with highest gain ratio were JobTitle, GSpecial, NKids, and Age, while the less effective attributes were MStatus, Uncomworkcond, UnivType, Grade and Dissatsalwork. The decision tree built using the C4.5 algorithm was so much pruned to consist of only three attributes; with PrevCo as the starting node, and Dissatsalrank and Grade as other attributes.

For the PrevCo attribute, it can be noticed that the employee performance varies from exceed and far exceed if the PrevCo is less than 3, then becomes accomplish, and raises again when the PrevCo more than 3.

In addition, when dissatsalrank is Yes, the employee performs was Accomplish, while the No value indicates the satisfaction of an employee to perform Exceed. This indicates a normal reaction of dissatisfaction in the salary or rank.

Grade attribute has an interesting result, which indicates that employees with grade good, is far exceed in the performance, while other grades are only exceed. This could be

indicating that graduates with high grades do not necessarily indicate good productive employees.

As an example of the generated C4.5 tree, Fig. 1 shows the tree generated for E2, and Table 4 shows the generated classification rules with the number of instances that support each rule.

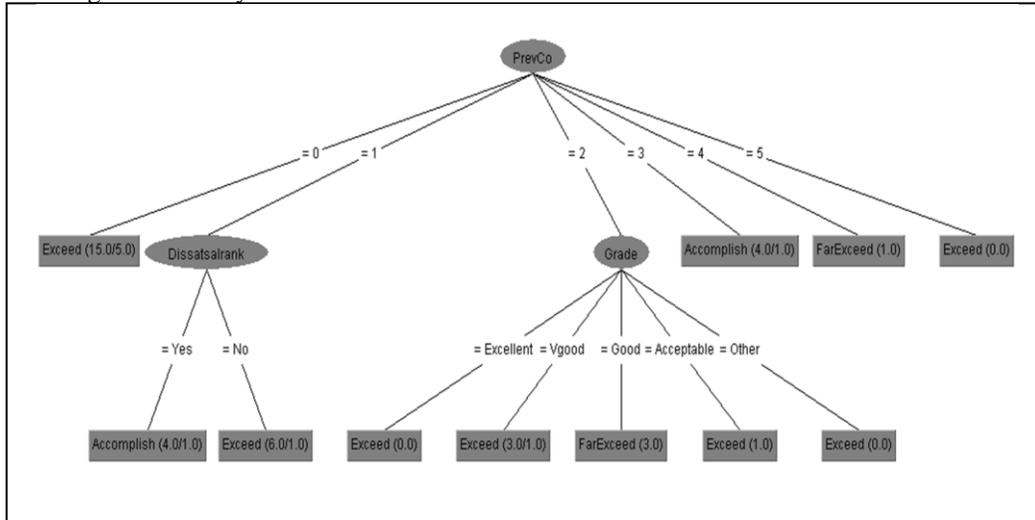


Figure 1. A decision tree generated by C4.5 algorithm for E2 for predicting performance

TABLE IV. CLASSIFICATION RULES GENERATED BY C4.5 ALGORITHM IN E2 FOR PREDICTING PERFORMANCE

Rule #	Rule Antecedent	Performance Decision	# of Instances
1	IF PrevCo = 0 ==>	Exceed	15
2	IF PrevCo = 1 & Dissatsalrank = Yes ==>	Accomplish	4
3	IF PrevCo = 1 & Dissatsalrank = No ==>	Exceed	6
4	IF PrevCo = 2 & Grade <> Good ==>	Exceed	4
5	IF PrevCo = 2 & Grade = Good ==>	FarExceed	3
6	IF PrevCo = 3 ==>	Accomplish	4
7	IF PrevCo = 4 ==>	FarExceed	1

C. Third Experiment (E3): Using the dataset gathered from the second IT company (38 instances)

Table 5 shows the accuracy percentages resulted from applying the algorithms of ID3, C4.5 and Naïve Bayes on the

dataset of the second IT Company. Note that the accuracy percentages were increased in this experiment.

The decision tree built using the ID3 algorithm for this experiment has started with JobTitle, as in E1. And then more weight has been given to attributes as GSpecial, Rank, Degree and ExpYears over attributes like SalRange, Grade and PrevCo.

TABLE V. ACCURACY PERCENTAGES FOR PREDICTING PERFORMANCE IN E3

Algorithm	10-Fold Cross Validation	Hold-out (60%)
ID3	50%	43.7%
C4.5 (J4.8)	60.5%	56.2%
Naïve Bayes	65.8%	68.7%

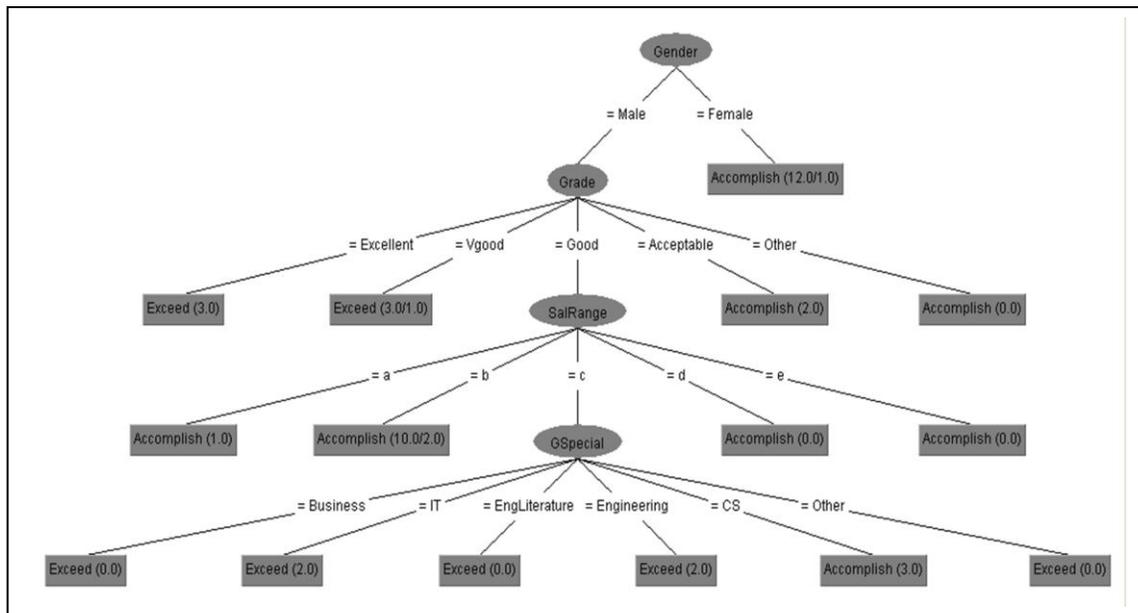


Figure 2. Decision Tree resulted from C4.5 algorithm for E3 to predict performance

TABLE VI. CLASSIFICATION RULES GENERATED BY C4.5 ALGORITHM IN E3 FOR PREDICTING PERFORMANCE

Rule #	Rule Antecedent	Performance Decision	# of Instances
1	IF Gender = Male & Grade = Excellent ==>	Exceed	3
2	IF Gender = Male & Grade = Vgood ==>	Exceed	3
3	IF Gender = Male & Grade = Good & SalRange = a ==> Performance =	Accomplish	1
4	IF Gender = Male & Grade = Good & SalRange = b ==> Performance =	Accomplish	10
5	IF Gender = Male & Grade = Good & SalRange = c & GSpecial = IT ==>	Exceed	2
6	IF Gender = Male & Grade = Good & SalRange = c & GSpecial = Engineering ==>	Exceed	2
7	IF Gender = Male & Grade = Good & SalRange = c & GSpecial = CS ==>	Accomplish	3
8	IF Gender = Male & Grade = Acceptable ==>	Accomplish	2
9	IF Gender = Female ==>	Accomplish	12

While using the C4.5 algorithm, the gain ratio of the Gender was the highest to start the tree with, and then comes Grade, SalRange and GSpecial. Fig. 2 shows the tree generated by WEKA for E3 and Table 6 shows the generated classification rules.

IV. RESULTS AND DISCUSSION

The study has found that several factors might have a great effect on employee performance. One of the most effective factors is the job title.

The trend of effectiveness of the job title is not much clear in the results, since there are about 20 job titles studied, but it

can be related to the type of job complexity and the responsibilities related to the title. High responsibilities sometimes affect the employee’s motivation and therefore performance in a positive way.

The university type attribute, in the three experiments, has positively affected the performance when the employee was graduated from a public university rather than a private one. This could be due to the fact that public universities accept, in most cases, students with high grades in high school comparing to private universities.

Other educational factors like degree and grade have slightly affected the performance, but not with clear trend, it might depend on other factors depending on the employee personality, which are not considered in this study.

Such, personality factors can be recognized by decision makers in interviews, so that they can complete their knowledge about the applicant. The university general specialization has a very close effect to performance as the job title. This could be due to the relationship between these two factors.

Some personal information like age, marital status and gender also affects the performance. Nevertheless, the age has not clear effect on the performance, since sometimes the performance increases with age, which adds the experience factor, other times, it decreases showing the highest motivation with the younger employees. Marital status, on the other hand, is clearer in its effect, since single employees in all experiments have shown better performance from married employees and even much better than married with kids employees.

But, surprisingly, in experiment E2, a strange trend appeared regarding number of kids, which indicated that the higher number of kids leads to a higher performance. This could be a coincidence outlier, since E2 dataset is not large enough to confirm this rule. Gender on the other hand, has no effect at all on experiments E1 and E2, since the female

proportion in both datasets is not significant. But in experiment E3, it indicated a higher performance for male employees than female.

Several professional factors also appeared to affect the performance. Salary, is one of the most positive factors on performance, this effects has been shown in experiments E1 and E3, while in E2 it was not significant. Number of previous companies in E1, showed both positive and negative relationship with the performance. This could be due to newly working employees who do not have experience working in other companies; they do their best to obtain better positions. On the other hand, employees worked in some previous companies may have much experience that would influence their performance. In E2 and E3 experiments only the positive relationship is observed. As for experience years, it affected the performance positively in E1 and E2, while in E3 it was not of much significance.

The Rank attribute has shown an interesting influence on performance, especially in E1 and E3. As in E2 it was not included as an effective factor. It was noticed in E1 and E3 experiments, the performance of senior employees is more than juniors. This is natural because of the experience of the employees that affect the Rank. But, surprisingly, team leaders and managers tend to have less performance than senior employees. This confirms the claims against highly positioned employees, that they do not work much.

Finally, job satisfaction and comfortable working environment has a slight effect on performance. For E3 they were not included as effective factors; while in E1 and E2 they were considered as low weighted effective factors on performance. This could be interpreted that the company in E3 has a more satisfactory conditions than the company in E2.

As a final remark on the accuracy of the classification models built for the three experiments, it can be noticed that for the different algorithms used, the classification accuracy was much more in experiments E2 and E3 than in E1. This might be because of the different companies of employees, included in E1, which created different factors affecting the classes in the experiments. While in E2 and E3, in spite of their small datasets, but the employees under study has the same working conditions, working environment, management and colleagues that made the study more focused on the measurable attributes on hand.

V. CONCLUSION AND FUTURE WORK

This paper has concentrated on the possibility of building a classification model for predicting the employees' performance. On working on performance, many attributes have been tested, and some of them are found effective on the performance prediction. The job title was the strongest attribute, then the university type, with slight effect of degree and grade.

The age attribute did not show any clear effect while the marital status and gender have shown some effect in some of the experiments for predicting the performance. Salary, number of previous companies, experiment years and job satisfaction, each had a degree of effect on predicting the performance.

For companies managements and human resources departments, this model, or an enhanced one, can be used in predicting the newly applicant personnel performance. Several actions can be taken in this case to avoid any risk related to hiring poorly performed employee.

As future work, it is recommended to collect more proper data from several companies. Databases for current employees and even previous ones can be used, to have a correct performance rate for each one of them.

When the appropriate model is generated, software could be developed to be used by the HR including the rules generated for predicting performance of employees.

REFERENCES

- [1] Allen, M.W., Armstrong, D.J., Reid, M.F., Riemenschneider, C.K. (2009). "IT Employee Retention: Employee Expectations and Workplace Environments", SIGMIS-CPR'09, May 2009, Limerick, Ireland.
- [2] Al-Radaideh, Q. A., Al-Shawakfa, E.M., Al-Najjar, M.I. (2006). "Mining Student Data Using Decision Trees", International Arab Conference on Information Technology (ACIT 2006), Dec 2006, Jordan.
- [3] Chein, C., Chen, L. (2006) "Data mining to improve personnel selection and enhance human capital: A case study in high technology industry", Expert Systems with Applications, In Press.
- [4] Cho, S., Johanson, M.M., Guchait, P. (2009). "Employees intent to leave: A comparison of determinants of intent to leave versus intent to stay", International Journal of Hospitality Management, 28, pp374-381.
- [5] CRISP-DM. (2007). Cross Industry Standard Process for Data Mining: Process Model. <http://www.crisp-dm.org/process/index.htm>. Accessed 10th May 2007.
- [6] Delavari, N., PHON-AMNUAISUK S., (2008). Data Mining Application in Higher Learning
- [7] Dreher, G.F. (1982) "The Role of Performance in the Turnover Process", Academy of Management Journal, 25(1), pp. 137-147.
- [8] Han, J., Kamber, M., Jian P. (2011). Data Mining Concepts and Techniques. San Francisco, CA: Morgan Kaufmann Publishers.
- [9] Ibrahim, M.E., Al-Sejini, S., Al-Qassimi, O.A. (2004). "Job Satisfaction and Performance of Government Employees in UAE", Journal of Management Research, 4(1), pp. 1-12.
- [10] Jantan, H., Hamdan, A.R. and Othman, Z.A. (2010a) "Knowledge Discovery Techniques for Talent Forecasting in Human Resource Application", International Journal of Humanities and Social Science, 5(11), pp. 694-702.
- [11] Jantan, H., Hamdan, A.R. and Othman, Z.A. (2010b). "Human Talent Prediction in HRM using C4.5 Classification Algorithm", International Journal on Computer Science and Engineering, 2(08-2010), pp. 2526-2534.
- [12] Karatepe, O.M., Uludag, O., Menevis, I., Hadzimehmedagic, L., Baddar, L. (2006). "The Effects of Selected Individual Characteristics on Frontline Employee Performance and Job Satisfaction", Tourism Management, 27 (2006), pp. 547-560.
- [13] Kayha, E. (2007). "The Effects of Job Characteristics and Working Conditions on Job Performance", International Journal of Industrial Ergonomics, In Press.
- [14] Khilji, S., Wang, X. (2006). "New evidence in an old debate: Investigating the relationship between HR satisfaction and turnover", International Business Review, In Press.
- [15] Lavrac, N. (1999). "Selected Techniques for Data Mining in Medicine", Artificial Intelligence in Medicine, 16(1999), pp. 3-23.
- [16] Mullins, I., Siadaty, M., Lyman, J., Scully, K., Garrett, C., Millar, W., Mullar, R., Robson, B., Apte, C., Weiss, S., Rigoutsos, I., Platt, D., Cohen, S., Knaus, W. (2006). "Data Mining and Clinical Data Repositories: Insights from 667,000 Patient Data Set", Computers in Biology and Medicine, 36, pp. 1351-1377.

- [17] Russ, F.A., McNeilly, K.M. (1995). "Links among Satisfaction, Commitment, and Turnover Intentions: The Moderating Effect of Experience, Gender and Performance", *Journal of Business Research*, 34, pp. 57-65.
- [18] Salleh, F., Dzulkifli, Z., Abdullah, W.A. and Yaakob, N. (2011). "The Effect of Motivation on Job Performance of State Government Employees in Malaysia", *International Journal of Humanities and Social Science*, 1(4), pp. 147-154.
- [19] Schwab, D. (1991). "Contextual variables in employee performance-turnover relationships", *Academy of Management Journal*, 34 (4), pp. 966-975.
- [20] Sexton, R.S., McMurtrey, S., Michalopoulos, J.O., Smith, A.M. (2005). "Employee Turnover: A Neural Network Solution", *Computer & Operations Research*, 32, pp. 2635-2651.
- [21] Witten I. Frank E., and Hall M. (2011). *Data Mining: Practical Machine Learning Tools and Techniques 3rd Edition*, Morgan Kaufmann Publishers.

AUTHORS PROFILE



Qasem A. Al-Radaideh is an Assistant Professor of Computer Information Systems at Yarmouk University. He received his Ph.D. in Data Mining field from the University Putra Malaysia in 2005. His research interests include: Data Mining and Knowledge Discovery in Database, Natural Language Processing, Arabic Language Computation, Information Retrieval, and Websites evaluation. He has several publications in the areas of Data Mining and Arabic Language Computation. He is currently the national advisor of Microsoft students partners program (MSPs) and the MS-Tech Clubs in Jordan.



Eman Al Nagi is a lecturer in the department of computer science at the Philadelphia University in Jordan. She received her MSc. In Information Technology Management from the University of Sunderland, UK on 2009. She has about 6 years experience in Software development and design