A COMPARISON OF PERFORMANCE PREDICTION IN THE COMPUTER PROFESSIONAL GROUPS BETWEEN MULTIPLE DISCRIMINANT ANALYSIS AND MULTINOMIAL LOGISTIC REGRESSION ANALYSIS

[1]NUTTHAPAT KAEWRATTANAPAT, [2]JARUMON NOOKHONG, [3]WANNARAT BUNCHONGKIEN, [4]MATHINEE KHONGSATID

Department of Social Sciences, Faculty of Humanities and Social Sciences, Suan Sunandha Rajabhat University
E-mail: nutthapat.ke@ssru.ac.th, jarumon.no@ssru.ac.th, wannarat.bu@ssru.ac.th, mathinee.kh@ssru.ac.th

Abstract- The purpose of this study was to create the equation and compare the accuracy of predicting to be the computer profession groups (Programmer, IT Support and System Analysis) in Bangkok Thailand. There are two statistical analysis in this article: Multiple Discriminant Analysis: MDA and Multinomial Logistic Regression Analysis: MLR. The psychology questionnaire by John L. Holland is used in this article for classifying 3 aspects: Data, Personal, and Tool. 89 persons, who work in computer field, participated in the study. The result of accuracy with 3 groups' computer professional showed that the overall accuracy is 66.7%. Furthermore, Programmer and IT Support has the overall accuracy is 93.1% and System Analyst is a high overlapping.

Keywords- Prediction, Computer Profession, Human Resource Management, Multivariate Analysis, Multiple Discriminant Analysis, and Multinomial Logistic Regression Analysis.

I. INTRODUCTION

Nowadays, there are so many unemployed person and many employed persons using internet for searching desired jobs and correspondent education level, as well as, human resource personnel of several organizations using such online media to be a medium for connecting to job finding person, which the labor behavior at the present time is to select jobs increasingly and websites attempts to inform job description for users. For the observation from pages of the websites and email sending job description to job applicants, the data is only recommended job description by bringing education qualification, graduated field and type of interested works in job searching in data base for recommending to the applicants only. The computer profession field is different, for example programmer, system analyst and IT support/ Help desk and etc. but recommending job description for applicants is not considered basic structure of a basic vocational orientation, which could reduce an opportunity for receiving work corresponding to ability and basic vocational orientation in job psychologically.

Therefore, the objective in the research is creating prediction equation and comparison of performance prediction in the computer professional groups from 2 types of analysis method, namely Multiple Discriminant Analysis: MDR and Multinomial Logistic Regression Analysis: MLR. For prediction in type of works in 3 computer professions and utilizing psychological questionnaire according to John L. Holland model is utilized for classifying 3 aspects: data, person, and tool by bringing to use with persons working in several computer fields, total 89 persons. The results of this research could be utilized in development of Career Path Recommended System or to be a tool in selection of applicants, including increasing chance in application corresponding to ability and basic vocational orientation increasingly.

II. LITERATURE REVIEW

For the research for comparing performance prediction in the computer professional groups between multiple discriminant analysis and multinomial logistic regression analysis to obtain data for analyzing discrimination by statistical model, the researchers, therefore, studied and searched research guideline data, including theory and technology related to the research as follows:

A. Classification Techniques

Classification techniques are widely used in data mining to classify data among various classes. Classification techniques are being used in different industry to easily identify the type and group to which a particular tuple belongs. Classification techniques used in prediction work , for example: Decision Tree, Support Vector Machine: SVM, Discriminant Analysis and Logistic Regression etc. To measure the accuracy of prediction methods, such as Accuracy or Mean Square Error: MSE etc., which mainly depends on the accuracy of data is compatible with more than one classification technique.

B. Multiple Discriminant Analysis: MDA

MDA is the analytical technique in statistics using in data group arrangement by bringing several independent variables in interval measuring standard or ratio to conduct format analysis to arrange group or

predict dependent variable or categorical variable or class label. The discrimination analysis has 2 types, that are Two-group Discriminant Analysis and Multiple Discriminant Analysis: MDA.

In analysis of Multiple Discriminant Analysis, the analyst must know a number of group and what data of independent variable in each set refers to. When the analysis is conducted, we will know equation demonstrating relation of the variables, which could be utilized in prediction to arrange group as the equation will stay in linear form as demonstrated in (1).

$$Y_i = a + b_{i1}X_1 + b_{i2}X_2 + \ldots + b_{ip}X_p \quad (1)$$

As

Y_i = Dependent variable or categorical variable or discriminant score.

b_{ip} = Coefficient of independent variable or weight of independent value.

X_p = Independent value.

a = Constant.

p = Number of independent value.

i = Number of dependent value or categorical variables.

C. Multinomial Logistic Regression Analysis: MLR

MLR is the analytical method in statistics which can demonstrate a relationship between independent variable and dependent variable and can tell that which independent variable influencing or affecting dependent variable. A logical regression analysis in 2 groups of dependent variables is binary logistic regression analysis. If there is an analysis of more than 2 groups in dependent variable, it is Multinomial Logistic Regression Analysis as the Multinomial Logistic Regression Analysis will utilize probability in discrimination as in (2) and (3).

$$Y_i = \frac{e^{u_i}}{\sum_{k=1}^{j} e^{u_k}} \quad (2)$$

$$u = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_n X_n \quad (3)$$

As

Y = Probability of occurrence in situation Y.

β_n = Coefficient of logistic regression.

e = Exponential Function.

i = Order of dependent variable.

j = All numbers of variables.

D. Classification Accuracy

Accuracy is the method used in comparison of performance prediction in the computer professional groups between multiple discriminant analysis and multinomial logistic regression analysis as demonstrated in (4)

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (4)$$

As

Accuracy = Value of accuracy.

TP = Value of positive true prediction.

TN = Value of negative true prediction.

FP = Value of positive false prediction.

FN = Value of negative false prediction.

E. Questionnaire for Data Storage

A testing tool to measure basic vocational orientation in computer professing which is the questionnaire test, total 18 choices could generate results as psychologically vocational orientation by giving score in 3 aspects, namely data, personal and tool.

1. You prepare a report to submit to the instructor.
   1.1. Searching data from library.
   1.2. Preparing with friends.
   1.3. Preparing by yourself, such as typing or printing with computer.

L means like/need (2 scores)
T means thinking/not sure (1 score)
NL means not like/not need (0 score)

No.	Personal: P Aptitude		Data: D Aptitude		Tool: T Aptitude	
1	1.1	L/T/NL	1.2	L/T/NL	1.3	L/T/NL
2	2.1	L/T/NL	2.2	L/T/NL	2.3	L/T/NL
3	3.1	L/T/NL	3.2	L/T/NL	3.3	L/T/NL

- Total scores in each type (Vertical) must not be more than 36 scores.
- Total scores (all 3 types) must be equal to 54 scores.
- Total scores in each choice (Horizon) must be equal to 3 scores.

Fig. 1 Example of basic vocational orientation questionnaire test.

III. RESEARCH METHODOLOGY

A research methodology to compare performance prediction in the computer professional groups between multiple discriminant analysis and multinomial logistic regression analysis has process in research as follows:

1) The testing tool of vocational orientation in computer professional was prepared.
2) Test and score collection was conducted by bringing the questionnaire to ask and collect score from the sample group (99 persons) as all of them are working in computer profession and after asking and obtaining the questionnaire back, total 87 persons, the obtain data was divided into programmer, total 29 persons, system analyst, total 29 persons and IT Support, total 29 persons.

3) Data preparation for data processing in the analysis was utilized the data form total 4 variables, divided in to 3 independent variables and 1 dependent variable as Table I

TABLE I
VARIABLES USED IN THE RESEARCH

Variables	Data Type	Variable Type	Description
Data	Numeric	Independent	Score of ability in Data
Personal	Numeric	Independent	Score of ability in Personal
Tool	Numeric	Independent	Score of ability in Tool
Position	Nominal	Independent	1 substituted Programmer, 2 substituted System Analyst and 3 substituted It Support

4) The analytical test in data and adjustment by Multiple Discriminant Analysis by dividing the test in 2 types:

Type 1: The test was conducted in every profession (Programmer, System Analyst and IT Support) and observed for obtaining results. If the predicted results have error in high level or overlapping in profession, the test may be conducted in type 2 for observing results. Type 2: The test was conducted in specific profession which did not have much overlapping and observed for obtaining results to obtain model used in prediction in other form.

5) The analytical test in data and adjustment by Multinomial Logistic Regression Analysis by dividing the test in 2 types in the same as step 4 as aforementioned.
6) The test results was compared and summarized.

IV. RESEARCH RESULTS

The test results of comparison of performance prediction in the computer professional groups between multiple discriminant analysis and multinomial logistic Regression Analysis could be summarized as follows:

A. The results of Multiple Discriminant Analysis

The results of Multiple Discriminant Analysis by applying Enter Independent Together with 3 professions have 66.7% accuracy which could be noted that the accuracy in prediction in programmer profession is equal to 89.7% and IT Support is equal to 82.8%, as for system analyst is equal to 27.6% which has low accuracy that could mean the system analyst have to possess several skills rendering an overlapping in skills. The accuracy in predicted group membership is demonstrated in Table II.

TABLE II
ACCURACY OF MULTIPLE DISCRIMINANT ANALYSIS IN 3 PROFESSIONS

Position	Predicted Group Membership			Total
	Programmer	System Analyst	IT Support	
Programmer	89.7%	0	10.3%	100%
System Analyst	24.1%	27.6%	48.3%	100%
IT Support	0	17.2%	82.8%	100%

66.7% of original grouped case correctly classified.

TABLE III
CLASSIFICATION FUNCTION COEFFICIENTS IN 3 PROFESSIONS

Skill	Positions		
	Programmer	System Analyst	IT Support
Data Skill	636.575	637.568	635.798
Personal Skill	632.106	633.428	631.918
Tool Skill	625.388	626.637	624.980
Constant	-17058.215	-17122.241	-17033.977

From the Table III, it tells that the function model of discriminant comprised coefficient and constant. Such discriminant and constant could be brought to build linear equations as follows:

Programmer = -17058.215 + 636.575(Data Skill) + 632.106(Person Skill) + 625.388(Tool Skill) (5)

System Analyst = -17122.241 + 637.568(Data Skill) + 633.428(Person Skill) + 626.637(Tool Skill) (6)

IT Support = -17033.977 + 635.798(Data Skill) + 631.918(Person Skill) + 624.980(Tool Skill) (7)

From such linear equations, they could be used in prediction for group arrangement when variable is substituted with Data Skill, Person Skill and Tool Skill if any equation that give the highest calculation value could be considered that such data set is arranged in professional group.

The results of Multiple Discriminant Analysis by utilizing Enter Independent Together with 2 professions are programmer and IT Support by excluding system analyst as there is an overlapping in skills, making high error. Therefore, the test was conducted with specific data of 2 professions, giving the results having 93.1% accuracy which could be noted that the accuracy in prediction in programmer profession is equal to 89.7% and IT Support is equal to 96.6% which has higher accuracy but obtains lower profession for prediction as this could be utilize in

profession prediction in programmer and IT Support. The accuracy results are as demonstrated in Table IV:

TABLE IV
ACCURACY OF MULTIPLE DISCRIMINANT ANALYSIS WITH PROGRAMMER AND IT SUPPORT

Position	Predicted Group Membership		Total
	Programmer	IT Support	
Programmer	89.7%	10.3%	100%
IT Support	3.4%	96.6%	100%

93.1% of original grouped case correctly classified.

TABLE V
CLASSIFICATION FUNCTION COEFFICIENTS WITH PROGRAMMER AND IT SUPPORT

Skill	Positions	
	Programmer	IT Support
Data Skill	4.597	3.797
Personal Skill	2.192	2.339
Constant	-66.452	-55.206

From the Table V: it tells that the function model of discriminant comprised coefficient and constant. Such discriminant and constant could be brought to build linear equations as follows:

$$\text{Programmer} = -66.452 + 4.597(\text{Data Skill}) + 2.192(\text{Person Skill}) \quad (8)$$

$$\text{IT Support} = -55.206 + 3.797(\text{Data Skill}) + 2.339(\text{Person Skill}) \quad (9)$$

B. The results of Multinomial Logistic Regression Analysis

Multinomial Logistic Regression Analysis with all 3 groups of profession. The analysis results are demonstrated in Table VI and VII:

TABLE VI
ACCURACY OF MULTINOMIAL LOGISTIC REGRESSION ANALYSIS IN 3 PROFESSIONS

Position	Predicted Group Membership			Percent Correct
	Programmer	System Analyst	IT Support	
Programmer	26	0	3	89.7%
System Analyst	7	10	12	34.5%
IT Support	0	7	22	75.9%
Overall %	37.9%	19.5%	42.5%	66.7%

66.7% of original grouped case correctly classified.

TABLE VII
CLASSIFICATION FUNCTION COEFFICIENTS IN 3 PROFESSIONS

Skill	Positions	
	Programmer	System Analyst
Data Skill	.749	6.155
Personal Skill	.167	5.829
Tool Skill	.385	6.013
Constant	-22.673	-323.086

The reference category is IT Support

From the Table VII: it tells that the function model of discriminant comprised coefficient and constant. Such discriminant and constant could be brought to build linear equations as follows:

$$\text{IT Support} = 0 \quad (10)$$

$$\text{Programmer} = -22.673 + .749(\text{Data Skill}) + .167(\text{Person Skill}) + .385(\text{Tool}) \quad (11)$$

$$\text{System Analyst} = -323.086 + 6.155(\text{Data Skill}) + 5.829(\text{Person Skill}) + 6.013(\text{Tool Skill}) \quad (12)$$

Multinomial Logistic Regression Analysis with 2 groups of profession (Programmer and IT Support) by analysis results in Table VIII and IX:

TABLE VIII
ACCURACY OF MULTINOMIAL LOGISTIC REGRESSION ANALYSIS WITH PROGRAMMER AND IT SUPPORT

Position	Predicted Group Membership		
	Programmer	IT Support	Percent Correct
Programmer	26	3	89.7%
IT Support	1	28	96.6%
Overall %	46.6%	53.4%	93.1%

93.1% of original grouped case correctly classified.

TABLE IX
CLASSIFICATION FUNCTION COEFFICIENTS WITH PROGRAMMER AND IT SUPPORT

Skill	Positions
	IT Support
Data Skill	-.622
Personal Skill	.129
Tool Skill	0
Constant	8.381

The reference category is Programmer.

From the Table IX: it tells that the function model of discriminant comprised coefficient and constant by utilizing only data and person variables, all of which could create linear equations as follow:

$$\text{Programmer} = 0 \quad (13)$$

$$\text{IT Support} = 8.381 + (-.622)(\text{Data Skill}) + (.129)(\text{Person Skill}) + (0)(\text{Tool Skill}) \quad (14)$$

C. The results of comparison of performance prediction between MDA and MLR

The results of comparison of performance prediction in the computer professional group between Multiple Discriminant Analysis and Multinomial Logistic Regression Analysis with data of 3 groups of profession as demonstrated in Table X and the specific analysis of 2 groups of profession as demonstrated in Table XI:

TABLE X
THE ACCURACY OF MULTIPLE DISCRIMINANT ANALYSIS AND MULTINOMIAL LOGISTIC REGRESSION ANALYSIS WITH 3 GROUP OF PROFESSION

Position	Predicted Group Membership						Total
	Programmer		System Analyst		IT Support		
	MDA	MLR	MDA	MLR	MDA	MLR	
Programmer	89.7%	89.7%	0	0	10.3%	10.3%	100%
System Analyst	24.1%	24.1%	27.6%	34.5%	48.3%	41.1%	100%
IT Support	0	0	17.2%	24.1%	82.8%	75.9%	100%

Overall Accuracy: 66.7%

TABLE XI
THE ACCURACY OF MULTIPLE DISCRIMINANT ANALYSIS AND MULTINOMIAL LOGISTIC REGRESSION ANALYSIS WITH SPECIFIC PROFESSION OF PROGRAMMER AND IT SUPPORT

Position	Predicted Group Membership				Total
	Programmer		IT Support		
	MDA	MLR	MDA	MLR	
Programmer	89.7%	89.7%	10.3%	10.3%	100%
IT Support	3.4%	3.4%	96.9%	96.9%	100%

Overall Accuracy: 93.1%

CONCLUSION AND DISCUSSION

The analysis to predict computer profession group has 66.7 accuracy which considers to be not high accuracy that may be because the professions of all 3 subject field have high relations and be in the same computer vocation; therefore, each skill may be highly overlapped and hard to discriminate with accuracy. In the prediction, it is necessary to monitor or collect data and analyze results consistently as age, tool and method of computer science may change swiftly, which could affect a change in behavior that would affect accuracy in prediction. A system may be developed to have data exchange with Web Service Technology from experts in the field to be automatic in prediction. In some case, the predicted type of works, which has probability in No.2 would be brought to utilize in consideration of data together with No.1 for increasing diversity of recommended professions. Human Resource Section may use this concept to analyze for arranging staff group and if any staff has psychology in professional orientation less than that should be; such staff may have additional training.

ACKNOWLEDGMENT

The authors would like to thank Suan Sunandha Rajabhat University, Bangkok, Thailand (http://www.ssru.ac.th/) to provide funding support to attend the dissemination of research on this and thank family, friends, colleagues and students in the field of Information Management for cooperation in research, all of you.

REFERENCES

[1] Kularbphettong, K. , Tongsiri, C. (2012). 'Mining Educational Data to Analyze the Student Motivation Behavior'. World Academy of Science, Engineering and Technology, International Science Index 68, 6(8), 1036 - 1040.

[2] Nutthapat Kaewrattanapat. (2009). '.NET Framework with WEKA'. journal of faculty of sciences and technology Suan Sunandha Rajabhat University, 1(1), 2009, pp.5-8.

[3] Lin, W. , Jou, Y. , Wu, Y. , Hsiao, Y. (2013). 'Data Mining Applied to the Predictive Model of Triage System in Emergency Department'. World Academy of Science, Engineering and Technology, International Science Index 78, 7(6), 1810 - 1817.

[4] Rai, B. (2014). 'A Study of Classification Models to Predict Drill-Bit Breakage Using Degradation Signals'. World Academy of Science, Engineering and Technology, International Science Index 92, International Journal of Social, Management, Economics and Business Engineering, 8(8), 2343 - 2346.

[5] Gary D. Gottfredson. (1999). 'John L. Holland's Contributions to Vocational Psychology: A Review and Evaluation'. Journal of Vocational Behavior 55, 15–40.

[6] Nutthapat Kaewrattanapat. (2010). 'Recruitment System based on SOAP and XML Web Services'. Human Resource National Conference, Chulalongkorn University.

[7] Nutthapat Kaewrattanapat., Wichuda kunnu. (2013). 'The Automatic Classification of Thai news by Similarity Method'.the 2th International Symposium on Business and Social Sciences: ISBSS, Osaka Japan.