X International Conference on Multidisciplinary Research
(IEI, Chandigarh) Institution of Engineers, India , Chandigarh
22nd February 2020 www.conferenceworld.in
ISBN : 978-81-944855-2-0

Employee Performance Predicting Using Classification Techniques

[1]Dr. J. Krishna

[1]Associate Professor of C.S.E, Annamacharya Institute of Technology,India

[2]D. Deekshitha, [3]P. Neelaveni, [4]S.Leelavathi, [5]V.Lakshmi Sai

[2,3,4,5]Student of C.S.E Department ,AITS, India

ABSTRACT:

Human Resources Management (HRM) has become one of the main interests of executives and decision-makers in almost every type of business to develop strategies for the proper discovery of highly qualified workers. Accordingly, management is involved in these employees ' results. Particularly to ensure that the right person is allocated at the right time to the convenient job. From here, there has been growing interest in the position of data mining (DM) that its aim is to discover information from huge amounts of data. In this paper, DM techniques were used to create a classification model to predict the output of employees using a specific dataset obtained from the Egyptian Civil Aviation Ministry (MOCA) through a questionnaire prepared and distributed for 145 jobs. Three key DM techniques have been used to develop the model of classification and to classify the most important variables that have a positive effect on performance. The methods are the Tree of Decision (DT), Naïve Bayes, and Vector Machine Support (SVM). In order to obtain a highly accurate model, several experiments were carried out based on the previous techniques used in the WEKA tool to allow decision-making, HR practitioners to anticipate and improve their employees ' efficiency.

Index Terms —Classification, C4.5 (J48), Data Mining, Employees' Performance, HRM, MOCA, Naïve Bayes, SVM

I .INTRODUCTION:

HRM has a leading role in determining productivity and performance to improve consistency. Organizations consider HRM as "practices of men." Therefore, it becomes the responsibility of the HRM to allocate the best employees to the appropriate job at the right time, train and qualify them, and build evaluation systems to monitor their performance and an attempt to preserve the potential talents of employees.

Knowledge can be derived using different methods, one of which is the use of DM technique. Classification is a predictive DM technique that uses proven results obtained from different data to predict data values. Classification technique is a supervised learning technique in DM and machine learning, while class or target class is known before. Construction of classification models from an input dataset is one of the most useful tasks in DM. Models

ISBN : 978-81-944855-2-0

are typically developed by the classification techniques used to predict future data trends. With classification, predictive models have the specific objective of enabling us to predict unknown variables values depending on previously known interest values of other variables.

The technique of C4.5 (J48) is one of the DT family. It can create both the decision tree and its sets of rules. In addition, it builds the tree for enhancing the prediction accuracy. Besides that, the models that are generated from the C4.5 (J48) are easily understandable because the extracted rules from the technique have a very explicit uncomplicated interpretation and has the advantage that does not need any field learning or parameter setting. Where, on the expected target, the researcher can easily detect the most important variables. J48 is the best implementation for the C4.5 rev. 8 technique and will be used in this analysis as a variant of the WEKA toolkit.

Another classification technique used to predict a target class is the Naïve Bayes classifier or the Bayesian theorem. In addition, it provides a unique approach to realizing different learning algorithms that do not explicitly use probabilities, depending on the probabilities in its calculations. Therefore, the results of this classifier are more accurate, effective, and more sensitive to recent data inserted to the dataset.

SVM is considered to be one of the most efficient supervised techniques of machine learning with a straightforward structure and high classification capability. In addition, SVM is recognized as the appropriate technique for classification in machine learning and DM, especially on linear and non-linear decision margins where high model accuracy can be produced. SVM has many advantages as it does not have a limit on the number of attributes and relies on the kernel trick to construct the model through expert knowledge of the kernel adjustment problem. Minimal Sequential Optimization (SMO) is an algorithm for SVM. It is known as an effective technique of classification to solve the optimization problem. SMO can be considered as the state–of–the–art approach in a non-linear SVM. SVM will train the dataset using SMO algorithm to build the prediction model.

II CONSTRUCTING THE CLASSIFICATION MODEL

For the purpose, which is constructing the classification model, the suggested approach has been adopted researching some variables that can influence and forecast the performance of the employees. To achieve this goal, a standardized guide is needed to develop a DM project lifecycle that includes certain steps including problem identification and task structuring, data collection and understanding, Data Preparing and Pre-processing, Modeling and Experiments, Testing and Evaluating.

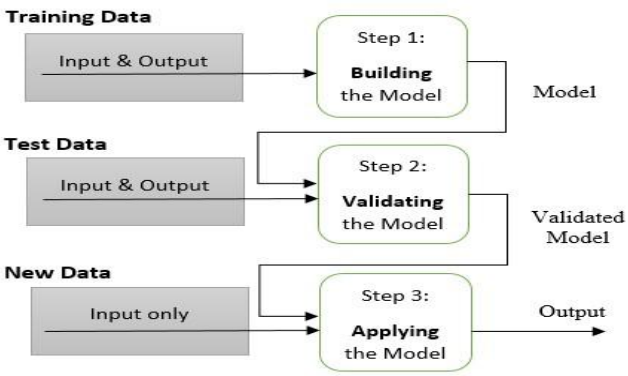

Fig.No:3. 0:The Classification Process in DM

ISBN : 978-81-944855-2-0

2.1. PROBLEM DEFINITION AND OBJECTIVE STRUCTURING:

This research focuses on how a proposed model supporting HRM and decision-makers can forecast the performance of MOCA employees and recognize the factors influencing and associating workers with poor / good performance. In addition,determining the most effective DM technique with the highest precision between the different classification techniques to be used.

2.2. DATA COLLECTION AND UNDERSTANDING PROCESS:

There needs to be a realistic way to collect the required data. A questionnaire is therefore prepared and circulated manually to MOCA staff, containing the various attributes that can influence and forecast the performance level (the goal class). The training dataset attributes requested are selected based on the relevant employee performance factors that are confined to Educational Factors, personal factors and factors such as (job title, age, rank, qualifications, grade, etc.) as shown in Table 1. Such characteristics are used to estimate the performance of the employee (the goal class) to be-excellent, decent, or good. 145 workers from all different sectors of MOCA completed the questionnaire with different job names, ages and grades to receive a full sample.

2.3. DATA PREPARATION AND PRE-PROCESSING:

The data processing process is carried out, with raw data comprising instances not relevant. This was due to errors and discarding anomalies. The data was transferred to Excel sheets in order to review and adjust the types of data collected where certain types of attributes need to be modified from numeric data type to categorical data type i.e. values shown by ranges for According to Table 1, the faculty specialization attribute (X15) included valuables such as IT, CS, MIS, which were considered to be a single value, IS, etc. Data generalization is therefore often known to be one of the methods of data reduction. After the sheet has been prepared and the processing needed The file was translated to an arff format compatible with the WEKA DM toolkit used in model construction.

The toolkit of WEKA (Waikato Environment for Knowledge Analysis) is a platform for machine learning developed by university researchers. Java is the language used for implementation. It offers a cohesive bundle in a single application that allows users to access new updated technology in the DM and machine learning world.

Consequently, the WEKA user can easily compare the results and accuracies of the implemented machine learning and DM algorithms in versatile procedures for a given dataset in order to detect the most suitable algorithm for the given dataset.

Variable Symbol	Variable	Description
A1	JT	Employee's Job Title
A2	Rank	Employee's Rank or Level
A3	Job Exp	No. of Working Experience Years

ISBN : 978-81-944855-2-0

A4	Duration	Service Period at MOCA (in Years)
A5	Prev Com	No. of Previous Companies the employee worked for
A6	Income	Range of Employee's Salary
A7	WorkCond	Working in Comfortable conditions (in employee's perspective). Answer with (Yes - No)
A8	Sat Income	Existing Satisfaction for Salary (in employee's perspective). Answer with (Yes - No)
A9	JobTrian	Existing trainings for the job (in employee's perspective). Answer with (Yes - No)
A10	Job Satis	Existing Satisfaction for the job (in employee's Perspective). Answer with (Yes - No)
A11	Age	Employee's Age
A12	Gender	Employee's Gender
A13	Mar Status	Employee's Marital Status
A14	Qualification	Employee's Educational Degree
A15	Special	General Specialization
A16	TU	Type of the University
A17	Grade	Employee's Graduation Grade
A18	Performance	Employee's Performance either as informed or predicted. This is the target class

III MODELING AND EXPERIMENTS

There have been three classification techniques that are classifier SVM, DT, and Naïve Bayes. Such classification techniques are used and applied to the data set to construct the performance prediction model of the employees in order to obtain the most suitable DM methodology and the most efficient variables that can influence and forecast the output of the employees. These variables consist of (A) Professional information such as: job title, rank, number of years of experience, number of years of service at MOCA, number of previously employed companies, salary, question about working in comfortable conditions, ask about the existence of comfort and satisfaction with salary, job, working conditions, and ask about training, (B) Personal information such as: age, (C) Academic information

ISBN : 978-81-944855-2-0

such as: grade, degree, general classification, and type of university. All these variables used to predict to be Excellent, Really Good OR GOOD the target class (MOCA employee performance).

3.0.1 First Experiment (E1): Using the whole variables of the dataset that may affect the performance (17 variables

No.	Technique	Prediction Accuracy
1	C4.5 (J48)	77.93 %
2	Naïve Bayes	71.03 %
3	SVM	81.38 %

Accuracy Percentages for Prediction Algorithms in E1

3.0.2 Second Experiment (E2): Using the important variables resulting from the use of Feature selection algorithms (10 variables)

Feature Selection Algorithm	Produced Feature Subset	Prediction Accuracy		
		Technique		
		C4.5(J48)	Naive Bayes	SVM
CORRELATION-ATTRIBUTEEVAL	[A2,A6, A9,A11, A10,A3]	79.31%	73.10%	*84.14%*

Accuracy Percentages for Prediction Algorithms in E2 Based On Using Feature Selection Algorithms

3.0.3 Experiment Three(E3): Using the most effective variables resulting from the tree generated using Decision Tree technique (5 variables)

No.	Technique	Prediction Accuracy
1	C4.5 (J48)	79.31 %
2	Naïve Bayes	82.07%
3	SVM	86.90 %

Accuracy Percentages for Prediction Algorithms in E3 Based On the Five Effective Variables

ISBN : 978-81-944855-2-0

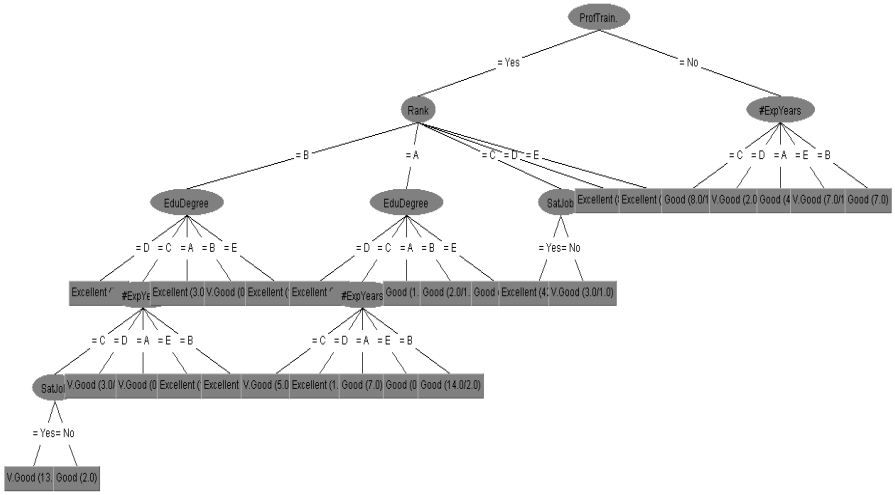

Fig.No:4.0.4The decision tree generated from using C4.5 algorithms for E3 to predict employees' performance

Rule #	Rule Antecedent	Performance Decision	# of Instances
1	If Job Train=yes & Rank = A & Qualification = D THEN	Excellent	7
2	If Job Train = yes & Rank = A & Qualification = C & #Job Exp = C EN	V.Good	5
3	If Job Train = yes & Rank = A & Qualification = C& #Job Exp = D EN	Excellent	1
4	If Job Train = yes & Rank = A & Qualification = C& #Job Exp = A EN	Good	7
5	If Job Train = yes & Rank = A & Qualification = C& #Job Exp = B EN	Good	14
6	If Job Train = yes & Rank = A & Qualification = A THEN	Good	1
7	If Job Train = yes & Rank = A & Qualification = B THEN	Good	2
8	If Job Train = yes & Rank = B & Qualification = D THEN	Excellent	2
9	If Job Train = yes & Rank = B & Qualification = C & #Exp = C &Job Satis = Yes THEN	V.Good	13
10	If Job Train = yes & Rank = B & Qualification = C & #Job Exp = C &Job Satis = NOTHEN	Good	2
11	If Job Train = yes & Rank = B & Qualification = C & #Job Exp = D THEN	V.Good	3
12	If Job Train = yes & Rank = B & Qualification = C & #Job Exp = E THEN	Excellent	1
13	If Job Train = yes & Rank = B & Qualification = C & #Job Exp =	Excellent	1

ISBN : 978-81-944855-2-0

	B THEN		
14	If Job Train = yes & Rank = B & Qualification = A THEN	Excellent	3
15	If Job Train = yes & Rank = B & Qualification = E THEN	Excellent	1
16	If Job Train = yes & Rank = C & Job Satis = Yes THEN	Excellent	42
17	If Job Train = yes & Rank = C & Job Satis = No THEN	V.Good	3
18	If Job Train = yes & Rank = D THEN	Excellent	8
19	If Job Train = yes & Rank = E THEN	Excellent	1
20	If Job Train = No & #Job Exp = C THEN	Good	8
21	If Job Train = No & #Job Exp = D THEN	V.Good	2
22	If Job Train = No & #Job Exp = A THEN	Good	4
23	If Job Train = No & #Job Exp = E THEN	V.Good	7
24	If Job Train = No & #Job Exp = B THEN	Good	7

IV RESULTS:

The purpose of this research was to detect the most suitable method of classification for the dataset used. Sequel to the above, model accuracy was used to define the dataset's most suitable classification technique. The model was created after a 10-fold validation technique was used to test the classification process. As shown in the following three tables (II, III and IV) Of the three experiments mentioned above, the SVM technique had the highest accuracy in all the experiments of the selected techniques. The SVM technique has been the most suitable classifier for the dataset as a result of the above.

As a final analysis of the accuracy of the classification models built by the three experiments, it was found that the accuracy of the prediction was much greater in E3 than in E2 and E1 experiments for all the different techniques used except the C4.5 (J48) technique. In E3 and E2 tests, it had the same precision, but it was much more than E1. This could be the less commonly used Variables in the classification process, the greater the accuracy of the classifier.

V CONCLUSION AND FUTURE SCOPE:

An important and urgent issue is the implementation of DM techniques in the various problem areas in the HRM field. Especially in Egypt's public sector. Therefore, to increase the horizons of academic and practice work on DM in HR to enter a high-performing government sector.

The SVM methodology was found to be the most appropriate classifier for the construction of the predictive model, where it had the highest predictive accuracy through all three tests with the highest percentage of 86.90%.

For decision-makers and HRM team, this or an improved model can be used to predict the performance of future talents that will be promoted, predicting the performance of recent candidate workers where different actions can be taken to avoid any risk associated with recruiting low-performance employees, or so on.

ISBN : 978-81-944855-2-0

As future work, to get high accuracy for the predictive model, it is suggested to help the used data set with a larger number of employees. To order to validate these results and help pick a more reliable model, the accuracy of other classification techniques such as Neural Network (NN), fuzzy logic and many others should also be experimented.

REFERENCES

[1] L. Sadath, (2013) "Data Mining: *A Tool for Knowledge Management in Human Resource*," International Journal of Innovative Technology and Exploring Engineering, *Vol. 2,* Issue *6, April 2013*.

[2] G. K. Gupta (2006) *"Introduction to Data Mining with Case Studies*" ISBN-81-203-3053-6.

[3] AI-Radaideh, Q. A., AI-Shawakfa, E.M., and AI-Najjar, M. I., (2006) "*Mining Student Data using Decision Trees*", International Arab Conference on Information Technology (ACIT'2006), *Yarmouk University, Jordan, 2006*.

[4] Surjeet K. Y., Brijesh B., Saurabh P., (2011) "Data Mining Applications*: A comparative Study for Predicting Student's performance*", International Journal of Innovative Technology and Creative Engineering, *Vol.1 No.12 (2011) 13-19*.

[5] Jantan, H., Hamdan, A. R., & 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 [D]*.

[6] Islam, M. J., Wu, Q. M. J., Ahmadi, M., and Sid-Ahmed, M. A., (2010), "*Investigating the Performance of Naive- Bayes Classifiers and K- Nearest Neighbour Classifiers"* Journal of Convergence Information Technology *Volume 5, Number 2, April 2010*.

[7] Al-Radaideh, Q.A., Al-Nagi, E., (2012). *"Using Data Mining Techniques to Build a Classification Model for Predicting Employees Performance",* International Journal of Advanced Computer Science and Applications*, 3(2), pp 144 – 151*.

[8] S. Yasodha and P. S. Prakash, (2012*), "Data Mining Classification Technique for Talent Management using SVM,"* the International Conference on Computing, Electronics and Electrical Technologies, *2012*.

[9] Hua Hu, Jing Ye, and Chunlai Chai, (2009), *"A Talent Classification Method Based on SVM",* in International Symposium on Intelligent Ubiquitous Computing and Education, Chengdu, China, *2009, pp. 160-163*.

[10] Kirimi JM, Motur CA (2016), *"Application of Data Mining Classification in Employee Performance Prediction"*. International Journal of Computer Applications, *Volume 146 – No.7, July 2016*.

[11] Desouki M. S., Al-Daher J (2015), *"Using Data Mining Tools to Improve the Performance Appraisal Procedure, HIAST Case".* International Journal of Advanced Information in Arts, Science & Management *Vol.2, No.1, February 2015*.

[12] V. Kalaivani, Mr.M. Elamparithi (2014), *"An Efficient Classification Algorithms for Employee Performance Prediction",* International Journal of Research in Advent Technology, *Vol.2, No.9, September 2014 E-ISSN: 2321-9637*.

ISBN : 978-81-944855-2-0

[13] Hamidah Jantan, Norazmah Mat Yusoff and Mohamad Rozuan Noh (2014), *"Towards Applying Support Vector Machine Algorithm in Employee Achievement Classification",* Proceedings of the International Conference on Data Mining, Internet Computing, and Big Data, Kuala Lumpur, Malaysia, *2014 ISBN: 978-1-941968-02-4 ©2014 SDIWC.*

[14] Lipsa Sadath (2013), "Data Mining: *A Tool for Knowledge Management in Human Resource",* International Journal of Innovative Technology and Exploring Engineering (IJITEE), *Vol-2, April 2013.*

[15] Qasem et al. (2012), *"Using Data Mining Techniques to Build a Classification Model for Predicting Employees Performance",* in (IJACSA) International Journal of Advanced Computer Science and Applications, *Vol. 3, No. 2, 2012.*

[16] 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.*

[17] Jantan, H., Hamdan, A.R. and Othman, Z.A. (2009), *"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.*

[18] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. Witten (2009), *"The WEKA data mining software: an update",* ACM SIGKDD Explorations Newsletter, *vol. 11, no. 1, pp. 10– 18, 2009.*