



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 10, October 2014

A Data Mining Model to Read and Classify Your Employees' Attitude

J John Victor Antony, G M Nasira

Senior Section Engineer, Indian Railways, Trichy, India

Assistant Professor, Chikkanna Government Arts College, Tirupur, India

ABSTRACT: We have learnt many lessons highlighting the fact that while the impact of employees with positive mind can surely put an organization on the trajectory of success, the presence of employees with negative attitude can not only diffuse the already existing positive air but can halt the growth engine altogether and push the organization into disastrously irrecoverable state. Hence, understanding the minds of the employees is of paramount importance for initiating proactive steps to sustain the growth momentum. The authors have evolved a model, explained in a step by step tutorial manner, to read and classify employees' attitude. The model employs K means clustering to construct an attitude reader and Rotation Forest to build an attitude classifier. Eventually, the model has been tested on a typical dataset pertaining to employees involved in train operations.

KEYWORDS: Clustering, Classification, Attitude Classifier, Attitude Reader, Railways

I. INTRODUCTION

Many organizational catastrophes are created by men than the machines that do not normally fail, with proven fail proof technologies in place. Executives of major corporations have often acknowledged the fact that the success and failure of a company greatly relies on the attitude of the people and the worth of the people surpasses even the biggest investment made on machines and materials. This is amply clear; when we look at the novel employee centric steps such as flexible work hours, work-at-home, sponsored get togethers, family travel trips the top companies all over the world adopt. These measures are directed at making the employees feel that the company is a great place to work. Despite the fact that a lot of innovative minds work together to chalk out strategies to secure employees' loyalty and establish positive work atmosphere, the study and management of human dynamics in an organization evades all approaches and continues to remain a formidable branch for study for ever, even with the availability of lot of analysis methods. As regards the application of technologies, we come across myriad situations in which artificial intelligence (AI) and data mining (DM) techniques are being used for fault pattern analysis through various sensor outputs, failure predictions, solution recommendations etc. Digging into medical parameters of patients, survival predictions and treatment recommendations are not uncommon. Applications into market analysis, customer purchase behaviour, bank loan clearance etc. and image processing are a series of areas the researchers frequent upon. All these areas have been visited and revisited by the scholars of computer science, each time trying to add a value in terms of algorithm, methodology and analysis. This apart, dealing with human resources in an organization which is a herculean task has, so far been the subject matter of social scientists and human resource managers. They resort to various analysis techniques, mostly qualitative in nature, the efficacy of which is an ongoing debate. The supremacy of qualitative versus quantitative methods and vice versa is also being continuously disputed, although the quantitative measurement of any parameter brings out an objective perspective, much to the convictions of many. Not inclined to get entangled in to the debate, the authors earnestly wish to present a DM model and say that the analysis of human dynamics can be strengthened and validated by adopting the rich techniques available in data mining analytics. An in depth literature survey highlights the fact that DM techniques can make greater inroads, filling many gaps, especially in analysis and thereby mutually reinforcing the interdisciplinary element between computer and social sciences. Through this research work, the authors specifically tend to evolve a model for study and analysis of human attitudes. The model, in its first stage, measures the parameters quantitatively using a well designed questionnaire, transforms the model suitably to qualitative one but carefully using numeric indices / labels with an aim of harnessing the power of DM algorithms, and



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 10, October 2014

progresses to analysis stage using clustering based attitude reader (AR) and finally ends up with the construction of an attitude classifier (AC).

II. BACKGROUND

A. Attitude: The Growth Driver:

It is a proven fact that every successful organization has effectively invested on the most valuable assets-the people [1]. They are indispensably precious assets whose value is more than the capital or equipment. The leaders of such organization have succeeded in creating an organizational culture that promotes positive attitude in its entire hierarchy. A positive attitude in an organization promotes

- Team work
- Reduction of wastes
- Increased productivity
- Securing employee loyalty

It is a matter of fact that today's biggest problem is people problem. The people, if properly nurtured, are the valuable assets and they become unbearable liabilities, otherwise. A Harvard university survey states that 80% of the times, the organizations recruit the people on the basis of their attitude. If such is the importance given to the attitude, the organizations need to read or assess the employees' attitude on a time bound manner, in order to ensure that the attitude level remains at its optimum and uncover potential issues even before they show up. Many factors go into shaping the attitude of employees. The following factors broadly determine the attitude level in an organizational setup.

- Participation of employees on decision making process
- The position or designation they hold (the nature of work)
- Intra organizational interactions
- Policy
- Technology
- Instruction flow
- Work environs
- Employee welfare
- Grievance redress mechanism

B. Problem Statement:

Given the importance of attitude and its potential on productivity, every organization tends to measure the same, with the help of human resource managers and/or social scientists, who most of the time, resort to qualitative assessment which are relative indicators of the ground truths and also render the output of the study less appropriate for pattern analysis. The top management will feel guided and strengthened if the outputs are provided in terms of quantitative figures and emerging patterns. The ground truth when presented in the form of quantitative distributions and patterns with the help of DM algorithms will equip the top management with evidence based insights and more objective perspective of the issues under study. The authors are primarily motivated to evolve a generic model that is more quantitative and appropriate for DM algorithms, for pattern generation and classification of the factors under study. A secondary motivation of the authors is to confine the parameters of the study specific to an organization involved in train operations. The organization is the largest employer having twelve lakh head counts, in the present context and is responsible for the train operations in the country. The job demands high level of coordination, planning, execution with many pressing constraints. The authors have chosen some of the job categories that handle sensitive operations and try to investigate their attitude level, using the proposed model. The nature of the job with regard to the employee categories chosen for study broadly involves

- Running trains.
- Inspection / Certification of carriages and wagons.
- Given the arduous nature, this calls for
- working at odd hours
- staying at out stations
- sustained mental concentration



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 10, October 2014

- physical fitness
- Inspecting at isolated long rugged stretches, pit lines etc.

These employees are known as safety categories of the organization. Despite slew of measures aimed at improving their job readiness and satisfaction through subject refresher training, yoga, stress management, various benefits and allowances, etc, their attitude seems to be low and calls for suitable intervention. It is a widely accepted fact that managing the employees who run the trains is tougher than the others and the gamut of issues concerning these employees is much pronounced and discussed at every managerial level and decision making process. Also the number of these employees forms sizable chunk of the entire work force of the organization. Although the management tackles the day to day issues by quick fix solutions, a systematic approach to study their attitude and analyse its pattern so that their problems are addressed on a long term basis, is the need of the hour.

- Can we measure/ read their attitudes objectively?
- Can we look for any attitude pattern?
- Can we classify their attitude so that proactive steps can be initiated?
- Can we have a data mining based model more suitable for this kind of application?

These are the pertinent issues the authors have addressed in this article. In this endeavour, they have attempted to derive a model for reading and classifying the employees' attitude.

III. RELATED WORK

While there have been many research efforts directed at studying and bettering the performance of machineries, a search to find relevant literature with regard to human resources, the understanding of which has always been a daunting and challenging task even now for social scientists that too, with the abundance of analytical tools, yields minimal results. More important, most of the studies are invariably bent on using methods such as variance, correlation and regression for analysis. Resorting to DM algorithms for study appears to be near non existent. The following brief account of relevant works by researchers bears evidence to this claim.

[2] addressed the issue of whether positive employee attitudes and behaviours influence business outcomes or vice versa with data pertaining to regional restaurant chain, using cross-lagged regression analyses. [3] showed that attitude towards pay is important for attracting employees while benefits are important for retaining them, using four years' data with assessment done each year to examine the temporal variations in the relationships over time. [4] have used machine learning techniques and a dictionary-based approach to detect comments expressing negative attitude of the users. The data used were user generated online comments on You Tube. [5] attempted to answer questions about job satisfaction and organizational commitment, by quantitative analysis of user questionnaire responses using hard C-means algorithm for vectors of fuzzy sets. [6] analysed employee attitude data from 35 companies over 8 years using lagged analyses and established a relationship between job satisfaction and organizational financial and market performance (return on assets/earning per share). The relationship between personality traits and their influence on job attitude has been studied by [7], using correlation and regression techniques. [8] used Theory of Reasoned Action (TRA), one of the social science theories, to understand knowledge sharing behavior and how extrinsic rewards can affect knowledge sharing attitude. [9] carried out correlation analysis and multiple stepwise regression analysis on survey quantified results by SPSS software and explored the mutual relationships among the three factors namely organizational commitment, employee satisfaction and work performance and gave recommendations to the management, so as to establish better organizational atmosphere, improve the staffs' work attitude and enhance their job performance. [10] tried to find the relationship between organizational politics, job satisfaction and organizational commitment, using three popular instruments such as Perception of Organizational Politics Scale, Job Satisfaction Survey and Organization Commitment Scale. [11] recorded significant attitude changes by subjecting the participants to a class room simulation exercise that included answering a well constructed questionnaire. [12] in their study, used a path model to analyze the so called three dimensions namely person to policy fit, person to supervisor fit and person to coworker fit, and showed that overall person to organization fit was positively correlated to job satisfaction, and person to supervisor fit had a direct effect on turnover attitude. [13] has taken workforce diversity in an organization as a major objective and estimated employee attitudes toward diversity and compared the same with other organizations by using survey method and statistical analysis. [14] analyzed the relationship among some basic demographic characteristics of the employees such as their age, gender, level of education, work experience, EI levels, and their work attitudes using questionnaire survey data from eight different enterprises and statistical analysis. [15] worked on

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 10, October 2014

predictor variables of college students' attitudes toward seeking professional psychological help, using regression and variance analysis. [16] presented a simulation model, built on system dynamics methodology which is a mathematical modeling technique to understand the behaviour of complex systems over time. They focused on enabling disabled people to move from welfare to work by incorporating attitudes of disabled people and employers and bringing out key factors that enable and encourage disabled people to take up jobs.

IV. PROPOSED MODEL

A self explanatory pictorial representation of the proposed model is presented below. It shows the sequence of design of questionnaire, dataset, construction of AR and AC and the proactive steps an organization can initiate for improved organizational growth.

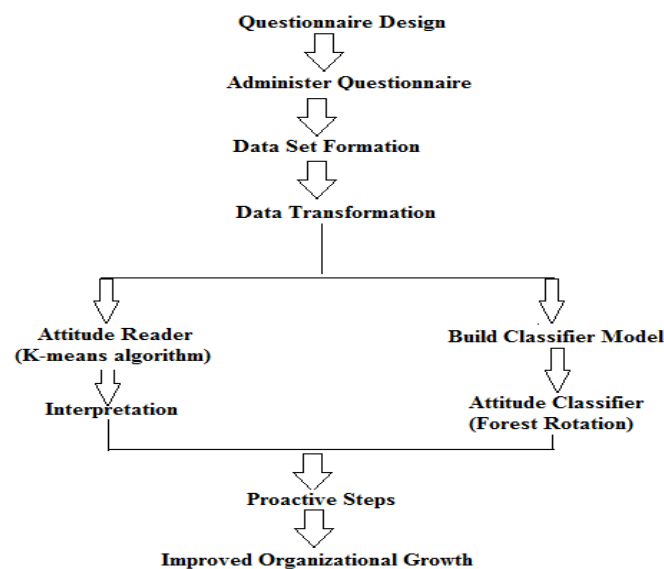


Figure 1: Proposed Model

A. Questionnaire:

A questionnaire containing twenty five questions was prepared. The questions which are objective in nature are directed at measuring the following attributes.

- Organizational commitment
- Satisfaction with work
- Satisfaction with pay
- Satisfaction with promotions
- Satisfaction with supervisor
- Satisfaction with job in general
- Intentions to quit

Each question was based on 2-point response scale with straight “Yes” or “No”. A “Yes” response would get the participant + 4 marks, while a “No” response, - 4 marks. The total score would be of + 100 or - 100 and is called as attitude score. Thus the score was kept numeric in nature with an aim of making the response score amenable to analysis.

B. Dataset:

The fields considered are presented in the Table 1. The fields were transformed so that DM tools can be applied. The transformed fields are provided in Table 2. It can be noted that the domain of “attitude” field of the table has four ranges that represent the intensity of attitude level. These four ranges were created by setting cut points using a free data pre-processing tool called Data Preparator (DP).

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 10, October 2014

Field	Type	Description
EmployeePF	Nom	Unique provident fund number of the employee
Department	Nom	Employee Department
Basic	Int	Basic Pay
Age	Int	Employee Age
Experience	Int	Years the employee served
EQ	Nom	Educational Qualification
Attitude	Int	Attitude Score
JobTitle	Nom	Position held by the employee

Table 1: Data Fields Considered

Field	Transformed Type	Transformation/Discretization
Basic	Nom	Discretized in to ranges of every 10000
Age	Nom	<25_30), <30_40), <40_50), <50_60>
Experience	Nom	<5_10), <10_20), <20_30), <30_40), <40_50>
EQ	Nom	Graduation, Non Graduation
Attitude	Nom	<0_50), <50_100), <-50_0), <-100_-50>
JobTitle	Nom	LPG, Sr ALP, ALP, LPM, App ALP, SSE/C&W

Table 2: Field

Transformations

V. MODEL REALIZATION AND DISCUSSION

A. Experimental Tool:

Weka 3.6.7 which is an Open Source Data Mining tool is considered for use. Written in Java and developed at the University of Waikato, New Zealand, it is freely available under GNU General Public License. Weka stands for Waikato Environment for Knowledge Analysis. Weka 3.6.7 is a collection of machine learning algorithms for data mining tasks [17]. It contains tools for Data pre processing, Classification, Regression, Clustering, Association rule mining, Feature Selection and Data visualization. [18] gives an account of Weka interfaces.

B. Attitude Reader:

AR has been realized through K-means algorithm which is a rigid clusterer. Rigid clustering refers to partitioning method in which a scheme called exclusive cluster separation is followed i.e. each data point belongs to exactly and only one of the partitions. K means algorithm adopts such a partitioning scheme. It uses either the default Euclidean distance or the Manhattan distance measures. If the Manhattan distance is used, then centroids are computed as the component wise median rather than mean. K-means algorithm is applicable, when the mean of a set of objects is defined but it is not possible to define means of nominal variables as we have transformed the data fields into ranges. In such scenario, it outputs for each cluster, the frequency counts of the values of each nominal attribute. This feature can be used to analyze nominal data by treating it as k way categorical variable with k different values or states and accordingly creating data summary by either using a coding scheme for all nominal attributes or otherwise [19]. To understand the ground truth available in the data set in terms of frequency counts, K-means algorithm was hence chosen and configured. The parameter configuration deals with setting the values for the important parameters of the data mining software during experimental run. The key parameter values are set as below.

- displayStdDevs = True . This setting sends the frequency counts of the nominal variables to the output which is the desired feature for the present context, otherwise it outputs standard deviations of numeric attributes
- distanceFunction = Euclidean distance or Manhattan distance
- number of clusters = 1. This is intentionally set so, as we are interested in generating the output of the entire data set as one entity (cluster).

The other values are allowed to be at their default setting. The output obtained is given in Table 3. The frequency distributions are provided in Fig 2, 3, 4 that are self explanatory, for better understanding of the users. The ground truth of the data set when presented in the form of frequency counts for the selected variables can better be visualized and

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 10, October 2014

understood by the users. It is iterated that K-means algorithm available in Weka is found to be efficient in generating the outputs in terms of frequency counts.

Attribute	Full Data (816)	Cluster #1(816)
Department		
Mechanical	816 (100%)	816 (100%)
Basic		
<10000_20000)	372 (45%)	372 (45%)
<0_10000)	307 (37%)	307 (37%)
<20000_30000)	137 (16%)	137 (16%)
Age		
<40_50)	163 (19%)	163 (19%)
<30_40)	386 (47%)	386 (47%)
<50_60>	130 (15%)	130 (15%)
<25_30)	137 (16%)	137 (16%)
Experience		
<10_20)	156 (19%)	156 (19%)
<5_10)	462 (56%)	462 (56%)
<20_30)	110 (13%)	110 (13%)
<30_40)	88 (10%)	88 (10%)
Attitude		
<-100_-50)	26 (3%)	26 (3%)
<-50_0)	280 (34%)	280 (34%)
<0_50)	368 (45%)	368 (45%)
<50_100>	142 (17%)	142 (17%)
Job Title		
LPG	205 (25%)	205 (25%)
Sr ALP	291 (35%)	291 (35%)
ALP	36 (4%)	36 (4%)
LPM	47 (5%)	47 (5%)
App ALP	83 (10%)	83 (10%)
SSE/C&W	154 (18%)	154 (18%)

Table 3: Attitude Reader Output

Fig 2 gives the frequency distribution of job titles available in the dataset. There were six job titles considered for this study.

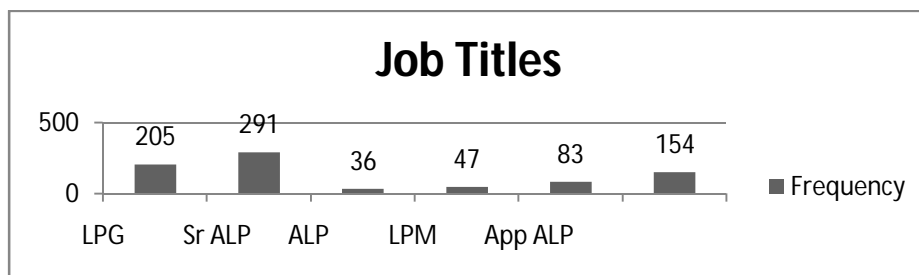


Figure 2: Job Title Distribution

Fig 3 gives the frequency distribution of the measured attitude of the employees. The “attitude” was consolidated into four ranges such as <-100_-50), <-50_0), <0_50) and <50_100>.

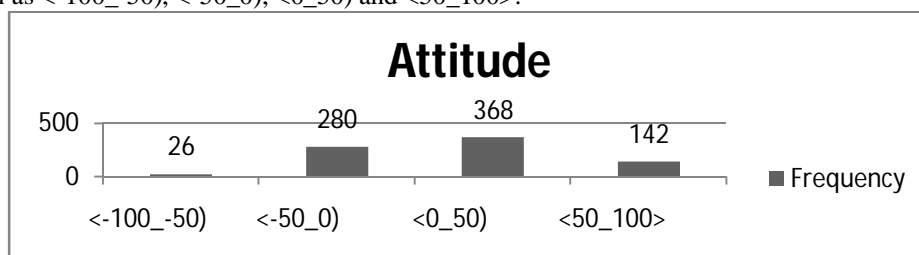


Figure 3: Attitude Distribution

Fig 4 gives the frequency distribution of the employees whose attitude falls in the range <-50_0). Similar such distributions for other attitude ranges can be generated.

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 10, October 2014

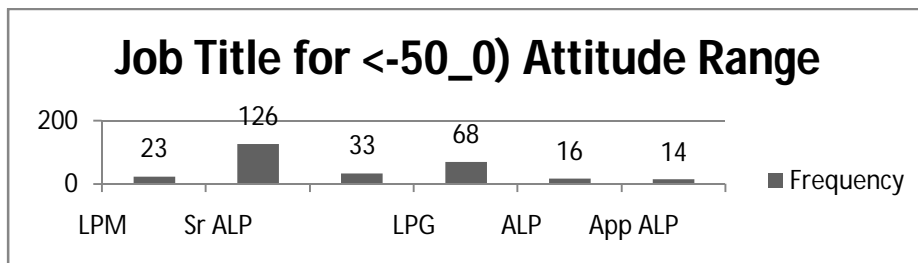


Figure 4: Distribution for Attitude Range <-50_0)

C. Attitude Classifier:

AC has been realized with the help of Classification algorithm which relies on inductive learning [20] in which a model is developed using sufficient number of known example cases from the data tuples, often, exploiting experts' knowledge. When class labels of the tuples are known, then, it is called as supervised learning. The set of tuples with known class labels is used for training the learning function. The model trained on this training set is capable of predicting future unseen cases using the previous knowledge. The aim here is to assign correct class to previously unseen and unlabeled nominal data objects. The original dataset was given as input to the classifier algorithm "Rotation Forest". The field "attitude" was set as class variable. It was found that the classification accuracy was 86%. In order to increase the classification accuracy, the data set was replicated so that sufficient tuples were made available for training the classifier model. After data replication with help of DP, the accuracy went up to 99%. The experiments were repeated using the other classifier algorithms. The results were found satisfactory. Rotation Forest algorithm with its proven efficiency and accuracy [21] was used to classifying the employees' attitudes. The algorithm's out put summary is provided below in Fig 5 and Fig 6.

Incorrectly Classified Instances	2	0.11%
Kappa statistic	0.9982	
Mean absolute error	0.0067	
Root mean squared error	0.0398	
Relative absolute error	2.0857	%
Root relative squared error	9.9262	%
Total Number of Instances	1765	
Ignored Class Unknown Instances	1	

Figure 5: Classifier Statistics

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	0.985	0	1	0.985	0.992	1	<-100_-50)
	1	0.001	0.998	1	0.999	1	<-50_0)
	0.999	0	1	0.999	0.999	1	<0_50)
	1	0.001	0.996	1	0.998	0.999	<50_100>
Weighted Avg.	0.999	0	0.999	0.999	0.999	1	

===Confusion Matrix ===

A	B	c	d	<--	classified as
65	1	0	0		A = <-100_-0)
0	605	0	0		B = <-50_0)
0	0	816	1		C = <0_50)
0	0	0	277		D = <50_100>

Figure 6: Classifier Accuracy



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 10, October 2014

VI. CONCLUSION

The main thrust of the article is to showcase the application suitability of DM methodologies for handling human resource related issues. A generic model using data mining has been presented and its viability verified using appropriate data set and algorithms. One part of the model has shown how to read attributes' values in the form of frequency counts, using clustering concept, resulting in the construction of Attitude Reader built on K means algorithm. Attitude Classifier, the other part of the model built on Rotation Forest, a classification algorithm, demonstrated how it can be used to classify human attitude for planning proactive steps, organization wide. The authors strongly feel that data mining methodologies can go beyond just machines and sensor outputs and find DM appropriate for dealing with human dynamics, in an organizational setup. They, through their model, stand convinced that social scientists can reap the benefits by employing data mining for analysis and pattern generation, in addition to the traditional statistical techniques such as variance, correlation and regression etc. Finally, the proposed model is easily implementable by common users also.

REFERENCES

1. Khera, S., "You can win", Published by Macmillan, 2000.
2. Koys, D. J., "The effects of employee satisfaction, organizational citizenship behavior, and turnover on organizational effectiveness: A unit-level, longitudinal study", *Personnel psychology*, Vol. 54(1), pp.101-114, 2001.
3. Carraher, S. M., "Turnover prediction using attitudes towards benefits, pay, and pay satisfaction among employees and entrepreneurs in Estonia, Latvia, and Lithuania". *Baltic Journal of Management*, Vol. 6(1), pp.25-52, 2011.
4. Kandias, M., Stavrou, V., Bozovic, N., Mitrou, L., and Gritzalis, D., "Can we trust this user? Predicting insider's attitude via YouTube usage profiling. In *Ubiquitous Intelligence and Computing*", 10th IEEE International Conference on Autonomic and Trusted Computing (UIC/ATC), pp. 347-354, 2013.
5. Auephanwiriyaikul, S., Keller, J. M., and Adrian, A., "Management questionnaire analysis through a linguistic hard C-means". 19th International Conference of the North American pp. 402-406, 2000.
6. Schneider, B., Hanges, P. J., Smith, D. B., and Salvaggio, A. N., "Which comes first: employee attitudes or organizational financial and market performance?", *Journal of Applied Psychology*, 88(5), pp.836, 2003.
7. Judge, T. A., Heller, D., and Mount, M. K., "Five-factor model of personality and job satisfaction: a meta-analysis", *Journal of applied psychology*, Vol. 87(3), pp.530, 2002.
8. Amin, A., Hassan, M. F., Ariffin, M. B., and Rehman, M., "Theoretical Framework of the Effect of Extrinsic Rewards on Individual's Attitude Towards Knowledge Sharing and the Role of Intrinsic Attributes", *ICCTD'09 International Conference In Computer Technology and Development*, Vol. 2, pp. 240-243, 2009.
9. Jing, C., and Xiao-hua, S., "The relationship research among organizational commitment, employee satisfaction and work performance", *ICMSE 2009 International Conference in Management Science and Engineering*, pp. 619-624, 2009.
10. Singh, K., "Relationship between organizational politics, job satisfaction and organizational commitment in UiTM Sarawak", *Business Engineering and Industrial Applications Colloquium (BEIAC)*, 2012 IEEE, pp. 377-382, 2012.
11. Schumacher, T. R., "Simulation and attitude change: Evolving the research model", *Proceedings of the Thirtieth Hawaii International Conference on System Sciences*, Vol. 2, pp. 648-654, 1997.
12. Chunjiang, Y., Honglan, W., and Ye, L., "The effects of person-organization fit on job satisfaction and turnover intention—The case of service industries", *IEEE International Conference on Business Management and Electronic Information (BMEI)*, Vol. 2, pp. 320-323, 2011
13. Pierce, A., Goodwin, S., and Baker, A., "Work force diversity: employee attitudes and perceptions in Innovation in Technology Management-The Key to Global Leadership", *PICMET'97: Portland International IEEE Conference on Management and Technology*, pp. 312, 1997.
14. Guan, X., Wu, W., and Hu, W., "Study on the relationship among demographic statistics characteristics, EI levels and job performance of the employee", *2nd IEEE International Conference on Artificial Intelligence, Management Science and Electronic Commerce (AIMSEC)*, pp. 6416-6419, 2011.
15. Zhi-hong, H., "Predictors of college students' attitudes toward seeking professional psychological help", *4th IEEE International Conference on Bioinformatics and Biomedical Engineering (iCBBE)*, pp. 1-4, 2010.
16. Abdelgawad, A. A., Snaprud, M. H., Krogstie, J., and Brynn, R., "Disabled People from Welfare to Jobs: A Decision Support Tool", *Sixth IEEE Asia Modelling Symposium (AMS)*, pp. 132-137, 2012.
17. Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., and Witten, I. H., "The WEKA data mining software: an update", *ACM SIGKDD explorations newsletter*, Vol. 11(1), pp.10-18, 2009.
18. Witten, I. H., and Frank, E., "Data Mining: Practical machine learning tools and techniques", Morgan Kaufmann, 2005.
19. Pedhazur, E. J., and Schmelkin, L. P., "Measurement, design, and analysis: An integrated approach", Psychology Press, 2013.
20. Poole, D. L., and Mackworth, A. K., "Artificial Intelligence: foundations of computational agents", Cambridge University Press, 2010.
21. Kuncheva, L. I., and Rodríguez, J. J., "An experimental study on rotation forest ensembles in Multiple Classifier Systems", Published by Springer Berlin Heidelberg, pp. 459-468, 2007.