IJCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Vol.-7, Issue-1, Jan 2019 E-ISSN: 2347-2693

Knowledge Discovery Techniques in Human Talent Management: "A Knowledge Discovery Databases Approach to Conquer Employee Attrition Problem Using Data Mining Techniques"

K. P. Tripathi[1*], Ashutosh Gaur[2]

[1,2]Bharati Vidyapeeth Deemed University, Pune, India

[*]*Corresponding Author: kptripathi06@gmail.com, Tel.: +91-94067-81075*

Available online at: www.ijcseonline.org

Accepted: 22/Jan/2019, Published: 31/Jan/2019

***Abstract*—** Talent resource management is one of the complex tasks for human resource professionals to assign the right person for the right place at the right time in the organization. The sustainability of suitable employee in an organization is very crucial these days. In this competitive age, employees are switching the organization on some gain but the organization suffers a lot. This paper is mainly concerned with the application of the knowledge discovery technique in human resource management, particularly in talent resource management to conquer the employee attrition and predicting the possible attrition in future.

***Keywords*—** Soft Computing, Fuzzy logic, Decision Tree, Talent Management, KDD, Knowledge Discovery.

I. INTRODUCTION

Knowledge Discovery Techniques are the techniques to extract of hidden information from large volumes of data, automatically. It is the process to identify the useful information / pattern from the historical / transactional data to provide some business value to the organization. These days, data is treated as the treasure and various tools are emerging for treasure hunt. It was only a decade which changes the rhythm of traditional business. A big paradigm change in the business when intellectuals proposed the importance of data dumps not for record obligations but using it to understand their clients better.

This is the biggest challenge to optimize the talent resource within the organization. It involves complex procedure to identify the right person at right place in the organization. Normally, it is complex process and depends on various factors like qualification, skill set, attitude of the person, nature of the job etc. so each factor has its own importance. This is the top challenge among the HR staff to select the appropriate talent from the existing human resource and maintain the sustainability in the organization. [1] [2]

Data mining techniques are the methods widely used to extract knowledge from the historical as well as transactional data. Data mining supports the KDD as given in [10]. There are various methods of classification task in data mining approach like neural network, decision tree, SVM, etc. [3] [11]. Decision tree is one of the popular methods used for classification and regression applications can provide the model with logical rules, used in the prediction [4] [7]. In this research paper, we considered the issues of attrition rate of the employee in the corporate.

The entire paper is divided into 5 Sections. The organization of the entire paper is as follows, Section I contains the introduction of Knowledge Discovery Techniques in Talent Forecasting, Section II contain the Talent Acquisition, Section III contain the methodology, Section IV contain the results and discussion and the Section V concludes research work with future directions.

II. TALENT ACQUISITION

The corporate world and other industries are seriously facing the issue of talent draining. Employees are developing their own hypothesis regarding the satisfaction level in the organization. Corporate world developed many strategies to evaluate the performance of their employees. The review of individual's performance based on his or her accountability is one of the tasks of the supervisors or managers. The set of some rules or processes usually termed as "employee performance review chart" is one of the tool to record the employee's contribution to the organization. The attributes of employee performance review chart may have 'satisfaction level', 'how long the employee associated with the organization', 'work load', 'project completed', 'issues faced in the organization' etc. We have collected data from open source repository available online to execute experiments in our research work.

Talent acquisition is one of the complex tasks of human resource management which results in the optimized use of

the resources in the organization. These days, data mining and machine learning techniques are used to explore the possibilities to get some useful results to support such complex issues in the business as well as in other areas [2] [3] [6] [12] [13]. The development of the decision-making model for the optimal use of the resources can be the key factor in the effectiveness of the overall administration of the organizations. In the multinational companies, the numbers of employees are huge with different background and skills. It is difficult to the HR department to utilize this human resource effectively and efficiently. Apart from that another big issue is to sustain them.

In this competitive era, employees are leaving organization for their personal benefit. To overcome this complex issue of mix and match, as well as attrition, we proposed KDD techniques to be used and support talent management and sustainability [7] [8]. The basic objective of this paper is to understand the pattern of attrition from the historical trend and predict the possible attrition of the employee in neat future. We propped the decision tree as the classification and predicting tool in this paper.

The output of this paper is the identification of the pattern of attrition and prediction of the possible attrition of the employee in near future. We processed the employee information for our experimental results but anonymized their personal information for the sake of the information security.

In the organization, the biggest issue is to identify the reasons of talent drain. The dataset under consideration is the part of the past assessment reports of the employees. The data set includes the information of the employee like dept., time he associated with the organization, work load, project / responsibility associated, satisfaction level, salary, etc. We cleansed the dataset using standard data cleansing methods and prepare our training dataset for the classification [10]. We considered decision tree for the classification task. The classification is based on the past annual performance data of the employees.

III. METHODOLOGY

This research paper adopted the knowledge discovery techniques to discover employees' performance patterns from the existing employees' performance assessment reports using the decision tree classifier. The selection of the techniques is based on the capability of the decision tree in the case of multiple independent variables, over the common classification techniques like neural networks for classification and prediction in data mining. Decision tree follows the 'divide-and–conquer' approach from a set of independent instances [4] [5] [9].

We proposed decision tree classifier to unhide hidden knowledge (pattern) in the form of logical rule to understand the reason of attrition in the organization and predicting the next possible employee who may leave the organization in future. We have applied soft-computing method, to induce human like intelligence in this complex and highly cognitive process of human resource management.

We proposed two major objectives in this paper:

1. To understand why the employees are leaving the organization; and
2. To predict, who will be the next So that the organization will be ready to take appropriate action to overcome on these issues. Therefore, we propose to work with the HR department to gather relevant data about the employees and to communicate the significant effect that could explain and predict employees' departure.

Following are the few of the data fields, used in this experiment work:

- Past Evaluation
- Level of satisfaction
- Project completed
- Avg. workload monthly (in hrs)
- Years associated with the organization
- Any accident
- Promotion held (if any, in last 5 years)
- Departments
- Current Salary earning
- Employee left or not (emp Status)

The objective of this paper is to identify the factors affecting attrition of the employee from the historical trends and develop a forecasting model to predict next possible talent drain in the organization. This study is based on the classification model using decision tree. This forecasting model will predict which valuable employee may leave in the future. There are roughly 15000 records in the dataset considered for this research. We used R language and R studio to conduct all the experiments of data exploration and classification.

The entire experimentation is conducted in two phases: Data exploration and Classification & Prediction. The first phase is about the data exploration, data cleansing and pre-processing done to explore data for better understanding. The data set is divided into training and testing sets. The second phase is to generate the classification rules using decision tree classifier.

IV. RESULTS AND DISCUSSION

The outcome of the first phase i.e. data exploration is very helpful to understand data better to go further for the classification and prediction. Table 1.1 is telling about the general trend of the data, the satisfaction level of the employees is 62.05%, 57% employees are satisfied with the existing evaluation system, 6 projects were completed on an average, etc. Figure 1 to Figure 10 showing the outcome of the exploratory analysis, are useful to generate initial useful insights from the data.

Emp_Satis	62.05
Eval_Proc	57%
Project	6
WorkLoad	302 Hrs
Time_Spent	3.9 Years
Issues	15%
Left_Org	24%
Promo	3%

Figure 1 is describing about the employees who left the organization in the past with their respective departments. It reveals that the attrition rate is maximum in sales department and least in management department.

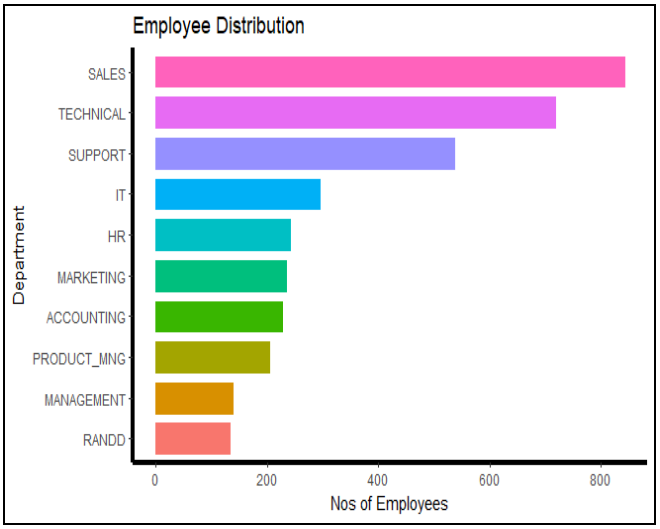

Figure 1: Employee left in the past with their respective department

Figure 2 is showing the comparison of the satisfaction level of the current and the previous employee who left the organization. It shows that the previous employee have high level of satisfaction who had 7.5 years or more of experience and least at 5.4 years of experience. The marketing department had the highest level of satisfaction in 7.5 year segment whereas the least in technical department in the same category. The current employees are not at that level even at 8 years of experience. Although the customer with an experience of 3 years, are more satisfied. Fig. 1.3 is displaying the correlation of all the numerical data points. It shows that the work load and the time were correlated with the decision of leaving. It means people who left the organization were overloaded. On average people who left, have a low satisfaction level, more experience in the organization, faced issues, they work more and didn't get promoted within the past five years.

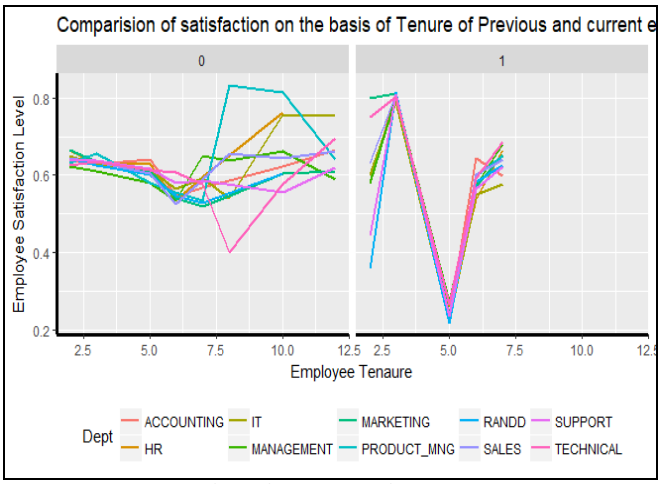

Figure 2: Comparison of Satisfaction level of the past and current employee

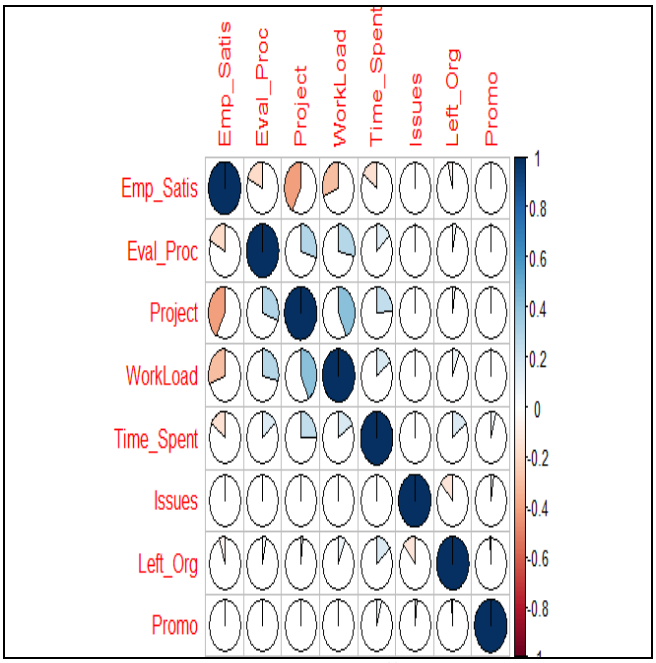

Figure 3: Correlation among different data points

Figure 4 is showing the satisfaction of the employees with number of projects of current and previous employees.

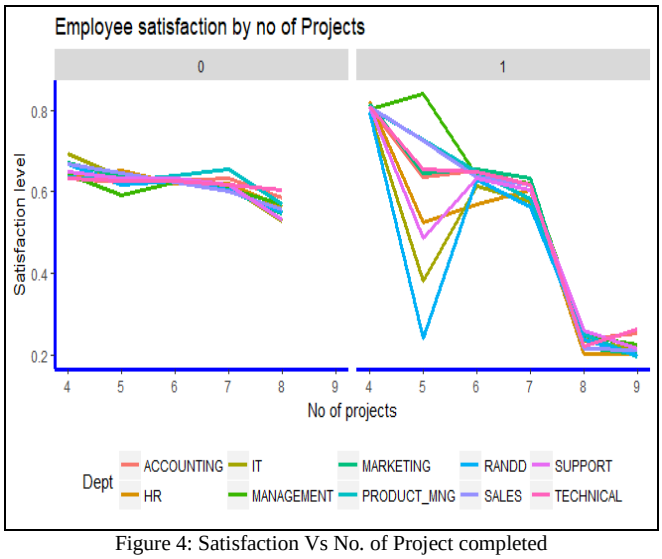

Figure 4: Satisfaction Vs No. of Project completed

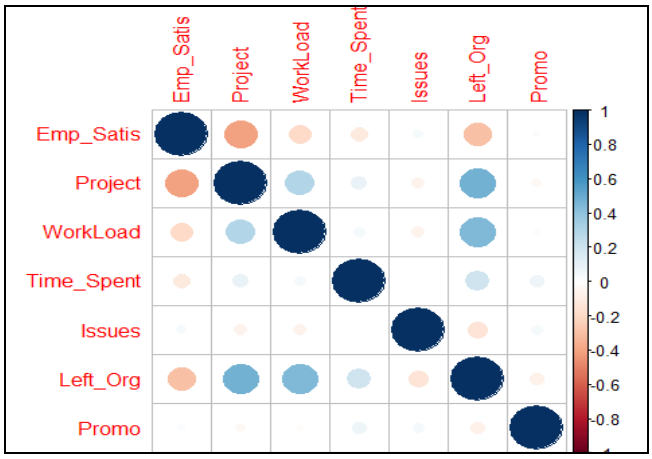

Figure 5: Correlation between Employee satisfaction with promotion and project

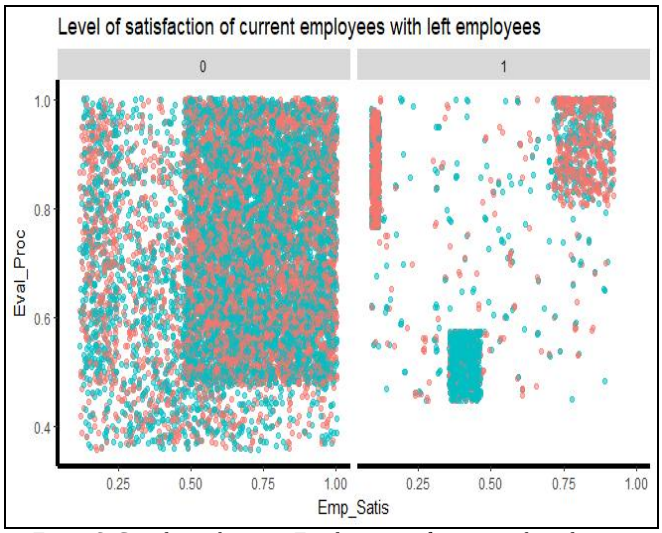

Figure 6: Correlation between Employee satisfactions with evaluation Process

Here it's much clearer from the Figure 5 that on average valuable employees that left the organization were not satisfied, they worked on many projects, spend many hours in the company but not promoted.

Fig.1.7 clearly tells that previous employees were more satisfied with more working hours as compare to present employees, who are not much satisfied with increase in work. One can easily identify three or four clusters in the figure1.7.

Decision tree: in decision analysis assignments, decision tree is the best tool to deal with the decision-making processes. In this paper, we used decision tree to derive the strategy to meet out our goal i.e. to identify the factors and the possibility of future attrition in the organization. Fig.1.8 is the decision tree on the given data. This decision tree was used for classification as well as prediction of the employees who left the organization or continued, on the basis of historical trend. The rule set of the decision tree is given in the fig. 1.9

Figure 7: Correlation between Employee satisfaction with work load

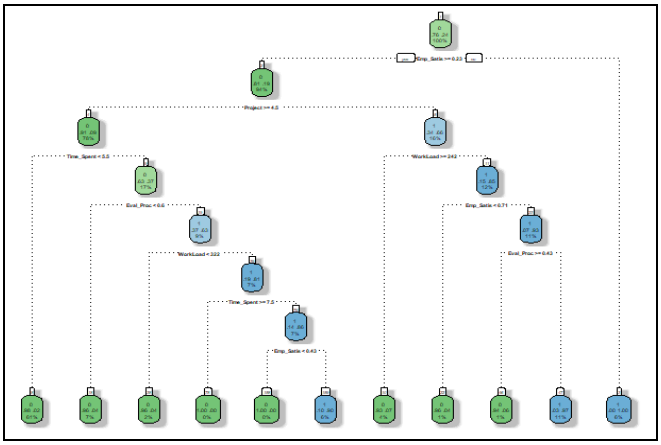

Figure 8: Decision Tree

n= 10499

node), split, n, loss, yval, (yprob) * denotes terminal node

1) root 10499 2512 0 (0.760739118 0.239260882)
2) Emp_Satis>=0.23 9895 1909 0 (0.807074280 0.192925720)
4) Project>=4.5 8181 772 0 (0.905635008 0.094364992)
8) Time_Spent< 5.5 6445 126 0 (0.980449961 0.019550039) *
9) Time_Spent>=5.5 1736 646 0 (0.627880184 0.372119816)
18) Eval_Proc< 0.605 761 33 0 (0.956636005 0.043363995) *
19) Eval_Proc>=0.605 975 362 1 (0.371282051 0.628717949)
38) WorkLoad< 322.5 228 9 0 (0.960526316 0.039473684) *
39) WorkLoad>=322.5 747 143 1 (0.191432396 0.808567604)
78) Time_Spent>=7.5 45 0 0 (1.000000000 0.000000000) *
79) Time_Spent< 7.5 702 98 1 (0.139601140 0.860398860)
158) Emp_Satis< 0.43 34 0 0 (1.000000000 0.000000000) *
159) Emp_Satis>=0.43 668 64 1 (0.095808383 0.904191617) *
5) Project< 4.5 1714 577 1 (0.336639440 0.663360560)
10) WorkLoad>=242 403 27 0 (0.933002481 0.066997519) *
11) WorkLoad< 242 1311 201 1 (0.153318078 0.846681922)
22) Emp_Satis< 0.71 122 5 0 (0.959016393 0.040983607) *
23) Emp_Satis>=0.71 1189 84 1 (0.070647603 0.929352397)
46) Eval_Proc>=0.435 54 3 0 (0.944444444 0.055555556) *
47) Eval_Proc< 0.435 1135 33 1 (0.029074890 0.970925110) *
3) Emp_Satis< 0.23 604 1 1 (0.001655629 0.998344371) *

Decision tree: in decision analysis assignments, decision tree is the best tool to deal with the decision-making processes. In this paper, we used decision tree to derive the strategy to meet out our goal i.e. to identify the factors and the possibility of future attrition in the organization. Fig.1.8 is the decision tree on the given data. This decision tree was used for classification as well as prediction of the employees who left the organization or continued, on the basis of historical trend. The rule set of the decision tree is given in the figure 9

Prediction

Prediction	0	1
0	3394	47
1	100	959

Figure 9: Decision Tree Rules

Accuracy = 0.9673333
Rules Generated by the decision tree for the prediction of attrition in the future.
n= 10499
node), split, n, loss, yval, (yprob) * denotes terminal node

1) root 10499 2512 0 (0.760739118 0.239260882)
2) Emp_Satis>=0.23 9895 1909 0 (0.807074280 0.192925720)
4) Project>=4.5 8181 772 0 (0.905635008 0.094364992)
8) Time_Spent< 5.5 6445 126 0 (0.980449961 0.019550039) *
9) Time_Spent>=5.5 1736 646 0 (0.627880184 0.372119816)
18) Eval_Proc< 0.605 761 33 0 (0.956636005 0.043363995) *
19) Eval_Proc>=0.605 975 362 1 (0.371282051 0.628717949)
38) WorkLoad< 322.5 228 9 0 (0.960526316 0.039473684) *
39) WorkLoad>=322.5 747 143 1 (0.191432396 0.808567604)
78) Time_Spent>=7.5 45 0 0 (1.000000000 0.000000000) *
79) Time_Spent< 7.5 702 98 1 (0.139601140 0.860398860)
158) Emp_Satis< 0.43 34 0 0 (1.000000000 0.000000000) *
159) Emp_Satis>=0.43 668 64 1 (0.095808383 0.904191617) *
5) Project< 4.5 1714 577 1 (0.336639440 0.663360560)
10) WorkLoad>=242 403 27 0 (0.933002481 0.066997519) *
11) WorkLoad< 242 1311 201 1 (0.153318078 0.846681922)
22) Emp_Satis< 0.71 122 5 0 (0.959016393 0.040983607) *
23) Emp_Satis>=0.71 1189 84 1 (0.070647603 0.929352397)
46) Eval_Proc>=0.435 54 3 0 (0.944444444 0.055555556) *
47) Eval_Proc< 0.435 1135 33 1 (0.029074890 0.970925110) *
3) Emp_Satis< 0.23 604 1 1 (0.001655629 0.998344371) *

We proposed decision tree as classifier in our study even other KDD tools are also available. The motive of considering decision tree was to keep analysis simple at the first stage. The 'goodness of fit' of a decision tree model can be evaluated by its ability to predict the outcome based on the previous training of the model using historical data [5]. Thus, decision tree follows a two-step process: first to train model on the historical data, and second to predict on the test data. One of the outcomes of the decision tree is the classification rule to predict the chances of attrition of an employee based on the performance of the employee, based on the various factors. We have identified some of the crucial factors which influenced the employee's decision to switch to other organization.

V. CONCLUSION AND FUTURE SCOPE

The motive of this research paper was to predict the employee attrition based on their performance score calculation using data mining techniques specially decision tree. This paper also described the significance of the data mining techniques in the prediction of the employee attrition rate using employee performance score. This Knowledge Discovery Databases based model can be implemented in the general Human Resource Management system to generate prediction on regular basis, so that HR manager can get an update about the possible loss of resource in advance and the best suitable action can be initiated to sustain the resource or to find best suitable match for the same.

This system would be useful for the higher and the middle level management for managing resources and get the optimum use of resources. The future scope of this paper could be implementation of another decision tree variants or other machine learning methods like neural network, fuzzy logic, SVM, hybrid methods etc. on the same problem.

REFERENCES

[1] A. S. Chang, & Leu, S.S., "Data mining model for identifying project profitablility variables," International Journal of Project Management, vol. 24, pp. 199-206, 2006.

[2] A TP Track Research Report "Talent Management: A State of the Art," Tower Perrin HR Services 2005.

[3] E. Frank, Hall, M., et al., "Data mining in bioinformatics using Weka," Bioinformatics Application Note, vol. 20, pp. 2479-2481, 2004.

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

[5] I. Bose, & Mahapatra, R.K., "Business data mining - a machine learning perspective " Information & Management, vol. 39, pp. 211-225, 2001.

[6] Jantan, H., Hamdan, A. R., & Othman, Z. A. (2010). Human talent prediction in HRM using C4. 5 classification algorithm, International Journal on Computer Science and Engineering, 2(08-2010), pp 2526-2534

[7] Jantan H., Hamdan A.R., Othman Z.A. (2009) Classification Techniques for Talent Forecasting in Human Resource Management. In: Huang R., Yang Q., Pei J., Gama J., Meng X., Li X. (eds) Advanced Data Mining and Applications. ADMA 2009. Lecture Notes in Computer Science, vol 5678. Springer, Berlin, Heidelberg

[8] J. Hamidah, H. Abdul Razak, and A. O. Zulaiha, "Classification for Talent Management using Decision Tree Induction Techniques," in 2nd Data Mining and Optimization Seminar (DMO'09), Bangi, Selangor, 2009, pp. 15-20.

[9] Kurgan, L.A., Musilek, P. (2006). A survey of knowledge discovery and Data Mining Models, The Knowledge Engineering Review, 21(1), pp 1 – 24

[10] Phyu, T.N., (2009). Survey of classification techniques in data mining, Proceedings of the International Multi Conference Of Engineers And Computer Scientists, IMECS 2009, Vol 1

[11] S. H. Liao, Chen, Y.N., & Tseng, Y.Y., "Mining demand chain knowledge of life insurance market for new product development," Expert Systems with Applications, vol. 36, pp. 9422-9437, 2009.

[12] V. Cho, & Ngai, E.W.T., "Data mining for selection of insurance sales agents," vol. 20, pp. 123-132, 2003

Authors Profile

K P Tripathi received his M.C.A. degree from Shivaji University, Kolhapur in June 2006. Presently he is pursuing his Ph.D. in Computer Application from Bharati Vidyapeeth Deemed University Pune, India. He has presented more than 10 papers in National International Conferences and published more than 20 papers in various National and International Journals. He is member of IACSIT, ICEIT and CSI. His areas of interest include Information Technology, Management Information System & Knowledge Management.

.

Ashutosh Gaur received Ph.D. degree in Artificial Neural Network. He has presented more than 20 papers in National and International Conferences and published more than 10 research papers in National and International Journals. He has authored more than 10 books on different topics published by Straight Forward Publication. He is a member of editorial board of various national and international journals. He is also a member of Computer Society of India. His areas of interest include E-Commerce, Artificial Intelligence and Artificial Neural Networks.

