



A Review of Factors and Data Mining Techniques for Employee Attrition and Retention in Industries

¹K. Mohammed Hussain, ²P. Sheik Abdul kadher

Research Scholar, Professor & Head of Department
Department of Computer Applications
B.S Abdur Rahman University, Chennai, Tamil Nadu, India
Email: Mohammed.hussain6@gmail.com

Abstract — In an exceedingly data driven economy, it's obvious that that raising as key competitive differentiators and retaining the talent pool has become a matter of dominant importance. The primary objective of this study is to provide a background on attrition and also to enlist various factors that build staff displeasure. This paper also provides inputs and reasoning on satisfactory level of staff towards their job and dealing conditions and to seek out the areas which are important for Indian IT industries. The secondary objective of this paper is to provide a detailed literature review of some of the suitable data mining techniques for employee attrition prediction and retention.

Keywords:- Data Mining, Employee Attrition, Prediction, HR

I. INTRODUCTION

Employee turnover refers to the proportion of employees who leave an organization over a period of one or two years, expressed as a percentage of total workforce numbers. This term is used to encompass all leavers, both voluntary and involuntary, including those who resign, retire or are made redundant, in which case it may be described as 'overall' or 'crude' employee turnover. It is also possible to calculate more specific breakdowns of turnover data, such as redundancy-related turnover or resignation levels, with the latter particularly useful for employers in assessing the effectiveness of people management in their organizations. Retention relates to the extent to which an employer retains its employees and may be measured as the proportion of employees with a specified length of service (typically one year or more) expressed as a percentage of overall workforce numbers.

Calculating your company's employee attrition rate allows you to determine the percentage of employees that left your business over a specified period of time, usually one year. Attrition includes all employees who leave the company, whether the leaving was voluntarily and involuntarily. An employee who chooses to leave a

company for another job is an example of voluntary employee attrition. On the other hand, an employee fired by the company is an example of involuntary attrition.

Both academic and industrial researchers start focusing on Employee retention and Employee voluntary turnover. Many of the researchers have examine the reasons for voluntary turnover and retention of professional sales force employees. Even fewer researchers have examined the effects of Human Resource Development (HRD) interferences on a sales force over an extended period of time.

One of the case study conducted for Industrial sector manufacturer, headquartered in India, examined the entire population of technical sales employees. The number of observations was extensive—over 20,000 observations associated with the 1,675 subjects analyzed for the study. The longitudinal period, size of the population, and the subject focus of this study distinguish this investigation from previously identified studies of employee voluntary turnover. The unique aspect of this study, however, lies in the number of variables and the variety of statistical treatments of employee turnover through the data-mining process.

One of the leading UK Annual Resourcing and talent planning survey report gives a median 'crude' or 'overall' employee turnover rate for the UK sample collected, as well as the median turnover figure relating purely to those who 'left voluntarily' (that is, resignations). While voluntary turnover rates have decreased recently as a result of challenging economic conditions, the flip side of this coin is that redundancy-related turnover has become more common. However, skills shortages persist for certain occupational groupings even during troubled economic times, so it is important to be aware of trends in turnover rates for different groups rather than simply focusing on 'headline' figures. Turnover levels can vary widely between occupations and industries. The highest levels are typically found in retailing, hotels, catering and leisure, call centres and

among other lower paid private sector services groups. Levels also vary from region to region. The highest turnover rates tend to be found where unemployment is lowest and where it is relatively easy for people to secure desirable alternative employment.

To know the acceptable level of employees towards their job and working circumstances, the organization should identify the factors which make disappointment of employees like policy or norms. Apart from this, organization should also find the areas where the company is lagging and also identify the reasons for attrition in Indian industries. Organization works towards methodologies and techniques to reduce attrition in the organization.

The aim of this paper is to check factors like remuneration, superior – subordinate relationship, growth chances, facilities, policies and procedures, recognition, gratitude, ideas, co-workers by that it helps to grasp the Attrition level within the organizations and factors about retain them. This study additionally provides a review of multiple papers on factors and issues related to employee's attrition and provide a detailed literature review on various researches conducted in employees' attrition prediction and application of data mining techniques.

II. LITERATURE REVIEW

Nagadevara et al, (2008), explored the relationship of withdrawal behaviors like lateness and absenteeism, job content, tenure and demographics on employee turnover in a rapidly growing sector like the Indian software industry. The sole aspect of this research was the application of predictive data mining techniques namely artificial neural networks, logistic regression, classification and regression trees, C5.0 classification trees and discriminant analysis). The authors worked on a sample data of 150 employees in a large software organization. The results of the study clearly illustrate a relationship between departure behaviors and employee turnover. This study also elevated several issues for future research. Further research works can be carried out to explicitly collect data on demographic variables across a large sample of organizations to assess the relationship between demographic variables and turnover. More analysis is recommended on large scale data on longitudinal mode on variables in the past academic research which have a relationship with turnover. Hamidah et al (2011), in their research paper detail the background of data mining, data mining in human resource application and also an overview of talent management. Based on the findings from the paper, there should be wider focus and research on different type of human resource applications and data mining techniques.

Jayanthi et al (2008) presented the role of data mining in Human Resource Management Systems (HRMS). This paper indicates that a deep sympathetic of the knowledge

concealed in Human Resource (HR) data is vital to a firm's competitive position and organizational decision making. Analyzing the patterns and relationships in human resource data is quite rare. The human resource data is usually treated to answer queries. Because human resource data primarily concerns transactional processing (getting data into the system, recording it for reporting purposes) it is necessary for HRMS to become more concerned with the computable data. They show how data mining discovers and extracts useful patterns from this huge data set to find noticeable patterns in human resource. The paper demonstrates the ability of data mining in improving the superiority of the decision-making process in HRMS and gives proposals regarding whether data-mining competences should lead to increased performance to survive competitive advantage. Wei-Chiang and Ruey-Ming (2007), in their work explored the feasibility of applying the Logit and Probit models, which have been effectively applied to solve nonlinear classification and regression problems, to employee voluntary turnover estimates. A numerical example involving voluntary turnover data of 150 professional employees drawn from a motor marketing enterprise in central Taiwan was used with a serviceable sample size of 132.

The data set was divided into two portions, the modeling data set and the testing data set using both logit and probit models. The testing data set was not used for either model building or selection, and was used for estimating model performance when applied to forthcoming data. The experimental results of their investigation exposed that the proposed models have high forecast capabilities and that the two (logit and probit) models also provide a promising alternative for predicting employee turnover in human resource management. The authors recommended that turnover research should move in new guidelines based on new expectations and methodologies, which would promote the new issues and problems. The authors proposed that neural networks and support vector machines can be used for classification problem for detecting the continuity of employees and identify who stays longer and who leaves sooner.

In a dissertation by Marjorie Laura Kane-Sellers (2007), the researchers carried out a study to explore the variables impacting employee voluntary turnover in the North American professional sales force of a Fortune 500 industrial manufacturing firm. By studying VTO (Voluntary Turn Over), the intention was to improvement a better considerate of HRD (Human Resource Development) interventions that could improve employee retention. The essential firm provided explanations of the employee database for all members of the professional technical sales force over a 14-years longitudinal period. The original database taken 21,271 discrete observations identified by unique employee clock number. The study design combined descriptive, correlation, factor analysis, multiple linear regression,

and logistic regression analysis techniques to examine relationships, as well as provide some predictive characteristics among the variables. Initially, evocative statistical techniques were used to develop baseline turnover rates, retention rates, and years of tenure. The mean tenure for the population as well as for each ethnic, gender, assignment location, supervisor, educational level, and sales training participation group was calculated. Hierarchical evocative techniques also provided the mean salary by job title, ethnicity, gender, educational level, and sales training participation. In this study, data-mining analysis started with descriptive analysis techniques.

The dynamic nature of Human Resource Management in ITES (BPO) sector has stimulated many researchers to study the various matters related to the high employee attrition in BPO industry. Anand et al. states that employee attrition discloses a company's internal power and faintness. Vijay and Sekar found that the research studies concentrating on capturing the perception of IT employees' knowledge about the ideal computer workstation arrangements and the optimal posture while working on computer is much limited in the literature. Mohamed et.al. observed that, from an organizational perspective, the higher the intra organizational trust, the more satisfied and productive the employees tend to be. New employee need to be continuously added, further costs in training them, getting them aligned to the company environment. Gupta reports that attrition is a burning problem for the promising industry of BPO, especially because it fails to tap the full utilization of the human resources and wastes much of its time, money and assets due to this[10]. Mike observed that Staff attrition (or turnover) represents significant costs to technology and business process outsourcing companies. High attrition rates drive up training costs, and increase human resources, recruiting, and output costs[11]. Khanna gives an overview of the BPO industry and analyzes as to how attrition is the predominant challenge facing the industry[12]. Agarwal feels that the challenge in the BPO industry is lack of discipline. BPO employees belong to a generation that does not like rules – they have had multiple choices from the time they were born, and the minute you hurt the dignity and self-respect of the people of this generation, they are bound to leave, which is probably the reason the attrition rate is so high, says Agarwal[13]. Kumar evaluated that the present salary package in BPO industry is not as lucrative as compared to other industries. Radhika observes that 40% attrition happens in first 120 days of hiring[14].

The importance of employee's retention and cost of employees' sendoff is well known in the literature. Resigning of an employee implies that employee is leaving with his or her implicit knowledge and thus it is a loss of social capital. Ongori, 2007 and Amah, 2009 indicated that attrition increases operation cost and cost on induction and training. The literature indicated that various factors that why employees quit job. There is also

much discussion on the relationship between various factors and attrition. For example, Mobley's (1977) study focused on the relationship between job satisfaction and attrition. Mohammad (2006) worked on the relationship between organization commitment and attrition. Another study to show the connection between work satisfaction, stress, and attrition in the Singapore workplace was conducted by Tan and Tiong (2006). Steijn and Voet (2009) also presented the relationship between supervisor and employee attitude in their study. A research was conducted in China to show the relationship between job satisfaction, organizational commitment or career commitment by Zhou, Long and Wang (2009). The results of each study were different as each study was carried out in different countries (having different socio-economic and culture), in different setting, for different organizations and used different independent variables.

III. IMPORTANT FACTORS

Review of various research studies indicated that employees quit for a variety of reasons, these can be classified into the following:

A. Demographic Factors

Various studies focus on the demographic factors to see attrition across the age, marital status, gender, number of children, education, experience, employment tenure.

B. Individual Factors

Individual factors such as health problem, family related issues, children education and social prestige contributes in attrition intentions. However, very little amount of empirical research work is available on individual related factors. There is another important variable "Job-Hoping" also donates in attrition intentions. Unrealistic expectation of employee is also an important individual factor which contributes in attrition. Several people keep impractical expectations from organization when they join. When these impractical expectations are not realized, the worker becomes disappointed and they quit. One of the individual features which have been lost in many research studies is the incapability of employee to follow organizations timings, rules, regulations, and requirement, as a result they resign.

Masahudu (2008) has identified another important variables "employers' geographic location" that may determine attrition. The intimacy of employees to their families and significant others may be a reason to look elsewhere for opportunities or stay with their current employers. For instance, two families living and working across two time zones may decide to look for opportunities closer to each other.

C. Propel factors

Drive factors are features that drive the employee towards the withdrawal from the employment. In the literature it is also called controlled factors because these factors are internal and can be controlled by organizations. According to Loquercio (2006) it is relatively uncommon for people to leave jobs in which they are happy, even when offered sophisticated pay elsewhere.

Most staff has a favorite for stability. However, some time employees are 'Propelled' due to disappointment in their present jobs to seek substitute employment. On the basis of available literature, Propel factor can be classified as follows:

D. Organizational Factors

There are many factors which are attached with an organization and work as Propel factors for employees to resign. Among them which are derived from various studies are: salary, benefits and services; scope of organization (the number of staff in the organization); location of the organization (small or big city); nature and sympathetic of organization; stability of organization; communication system in organization; management practice and policies; employees' authorization. There is another Propel variable called organizational justice.

According to Folger & Greenberg (1985), organizational justice means equality in the workplace. There are two methods of organizational justice: distributive justice, which describes the fairness of the consequences an employee receives; and procedural justice, which describes the equality of the procedures used to control those consequences.

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

Karatepe et al. (2006) defined the performance of a frontline employee, as his/her productivity comparing with his/her peers. Schwab (1991), on the other hand, defined the performance of university teachers included in this study, as the number of researches cited or published. In overall, concert is usually measured by the units produced by the employee in his/her job within the given period of time.

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

Previous studies stated several characteristics affecting the employee performance. Some of these attributes are personal characteristics, others are educational and lastly professional attributes were also measured. Chein and Chen (2006) used numerous attributes to imagine the employee performance. They specified age, gender, marital status, experience, education, major subjects and school tires as possible factors that strength disturbs the performance. Then they excluded age, gender and marital status, so that no discrimination would exist in the procedure of individual selection. As a result for their study, they found that employee recital is extremely pretentious by education degree, the school tire, and the job experience.

Kahya (2007) also searched on positive factors that disturb the job recital. The researcher reviewed earlier studies, describing the significance of experience, education, salary, working conditions and job satisfaction on the performance. As a outcome of the research, it has been create that several features pretentious the employee's performance. The position or grade of the employee in the company was of high positive result on his/her performance. Working circumstances and situation, on the other hand, had exposed both positive and negative relationship on performance. Highly educated and qualified employees showed disappointment of bad working conditions and thus pretentious their performance damagingly. Employees of low educations, on the other hand, showed high performance in malice of the evil conditions. In addition, experience showed positive relationship in maximum cases, while education did not yield strong relationship with the recital.

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

Jantan et al. (2010) had discussed in their paper Human Recourses (HR) system architecture to forecast an applicant's talent based on information filled in the human resource application and past experience, using Data Mining(DM) techniques. The goal of the paper was to find a way to flair guess in Malaysian higher institutions. So, they have specified certain features to be

considered as attributes of their system, such as, professional qualification, training and communal responsibility. Then, several data mining techniques (hybrid) were applied to find the prediction rules. ANN, Decision Tree and Rough Set Theory are examples of the selected techniques.

IV. CONCLUSION

Data Mining is an area full of exhilarating opportunities for researchers and practitioners. This field assists in industries with well-organized and effective ways to improve industrial effectiveness and employee efficiency. Data mining is a significant tool for helping organizations improve the decision making and analyzing new patterns and dealings among a huge amount of data. A broad sense of the types of research presently being lead in Data Mining was presented, from smearing data mining for understanding employee retention and attrition to finding new methods of making personalized learning recommendations to each individual employee. Many chances exist to study DM from an industrial unit of analysis to individual course-levels of analysis. Some effort is strategic in nature and some of the research is tremendously technical. A deep understanding of the facts and data hidden in Human Resource data is vital to a firm's competitive position and organizational decision making. Analyzing the patterns and relationships in Human Resource data is quite rare. The HR data is usually treated to answer queries. Because HR data primarily concerns transactional processing getting data into the system, recording it for reporting purposes it is necessary for Human Resource Management Systems to become more concerned with the quantifiable data. This paper discussed usefulness and application different mining techniques and useful factors for attrition prediction.

Multiple research avenues are available to improve the data mining to discover and extract useful patterns from the large data set to find observable patterns useful for effective attrition prediction. The focus on multi-dimensional hybrid decision tree based methodology would be helpful to improve the quality of the decision-making process in HRMS. More regression analysis and propositions on data-mining capabilities should be assessed to see if the mining methods can lead to high performance and accurate prediction of attrition for industries in India.

REFERENCES

- [1] Al-Radaideh, Q. A., Al-Shawakfa, E.M., Al-Najjar, M.I. (2006). "Mining Student Data Using Decision Trees", International Arab Conference on Information Technology (ACIT 2006), Dec 2006, Jordan.
- [2] Chein, C., Chen, L. (2006) "Data mining to improve personnel selection and enhance human capital: A case study in high technology industry", *Expert Systems with Applications*, In Press.
- [3] Cho, S., Johanson, M.M., Guchait, P. (2009). "Employees intent to leave: A comparison of determinants of intent to leave versus intent to stay", *International Journal of Hospitality Management*, 28, pp374-381.
- [4] CRISP-DM, (2007). Cross Industry Standard Process for Data Mining: Process Model. <http://www.crisp-dm.org/process/index.htm>. Accessed 10th May 2007.
- [5] Delavari, N., PHON-AMNUAISUK S., (2008). Data Mining Application in Higher Learning
- [6] Dreher, G.F. (1982) "The Role of Performance in the Turnover Process", *Academy of Management Journal*, 25(1), pp. 137-147.
- [7] Han, J., Kamber, M., Jian P. (2011). *Data Mining Concepts and Techniques*. San Francisco, CA: Morgan Kaufmann Publishers.
- [8] Ibrahim, M.E., Al-Sejini, S., Al-Qassimi, O.A. (2004). "Job Satisfaction and Performance of Government Employees in UAE", *Journal of Management Research*, 4(1), pp. 1-12.
- [9] Anand V.V., Saravanasudhan R. and Vijesh R., Employee attrition - A pragmatic study with reference to BPO Industry, In *Advances in Engineering, Science and Management (ICAESM)*, 2012 International Conference on Advances in Engineering, 42-48 IEEE (2012)
- [10] Vijay A. and Sekar A., New Quality Rating system for the Computer Workstation arrangements of the Information Technology Industries: A Six Sigma Model Approach, *Res. J. of Management Sciences*, 2(7), 15-21 (2013).
- [11] Mohamed M.S., Kader M. A. and Anisa H., Relationship among Organizational Commitment, Trust and Job Satisfaction: An Empirical Study in Banking Industry, *Res. J. of Management Sciences*, 1(2), 1-7 (2012).
- [12] Gupta S.S. Employee Attrition and Retention: Exploring the Dimensions in the urban centric BPO Industry", unpublished Doctoral Thesis, Retrieved from <http://www.jiit.ac.in/uploads/Ph.D-%20Santoshi%20Sen.pdf> (2010)
- [13] Mike. Employee attrition in India [Online Exclusive], *Sourcing Line*, Retrieved from

- <http://www.sourcingline.com/resources> accessed on January 7th, 2009 (2009)
- [14] Khanna R., How the BPO Industry has dealt with its Biggest Challenge: Attrition, online article accessed on November 2009. (2007).
- [15] Agarwal A., Challenge in the BPO industry is lack of discipline, Retrieved from www.docstoc.com/callcenter-kaleidoscope-events, accessed on 10th march 2009. (2008).
- [16] Kumar V., High Attrition rate attributed to pay package, Online Article Retrieved from <http://outsourcportfolio.com/high-attrition-rate-attributed-to-pay-package> on September 7, 2009 (2008).
- [17] Radhika. 40 percent attrition happening in 120days. Online article Retrieved from www.mckinsey.com/clients-service/bto/./pdf accessed on 23 rd February 2009, (2008)
- [18] Amah, O.E. (2008); Job Satisfaction and Attrition Intention Relationship: The Moderating Effect of Job Role Centrality and Life Satisfaction, Human Resources Institute & Curtin University of Technology, Singapore.
- [19] Barnard, M.E. and Rodgers, R.A. (1998); What's in the Package? Policies for the Internal Cultivation of Human Resources and for High Performance Operations, Asia Academy of Management (Hong Kong).
- [20] Bockerman, P. and Ilmakunnas, P. (2007); Job Disamenities, Job Satisfaction, Quit Intentions, and Actual Separations: Putting the Pieces Together, Discussion Paper No. 166, Helsinki Center of Economic Research, Finland.
- [21] Debrah, Y. (1993); Strategies for Coping with Employee Retention Problems in Small and Medium Enterprises (SMEs) in Singapore. Entrepreneurship, Innovation, and Change, 2, 2, pp 143-172.
- [22] Debrah, Y. (1994); Management of Operative Staff in a Labour-Scarce Economy: the Views of Human Resource Managers in the Hotel Industry in Singapore. Asia Pacific Journal of Human Resources, 32, 1, pp 41-60.
- [23] Folger, R. and Greenberg, J. (1985); Procedural justice: An interpretative analysis of personnel systems, Research in Personnel and Human Resources Management, 3: 141-183.
- [24] Johns, G. (1996); Organizational Behavior, New York: Harper Collins Publishing.
- [25] Loquercio, D. (2006); Attrition and Retention – A Summary on Current Literature, downloaded from People in Aid” <http://www.peopleinaid.org/> accessed on February 9, 2010.
- [26] Mobley and William. H. (1977); Intermediate Linkages in the Relationship between Job Satisfaction and Employee Attrition, Journal of Applied Psychology, Vol 62(2), April 1977, pp 237 - 240.
- [27] Mohammad et al, (2006); Affective Commitment and Intent to Quit: the Impact of Work and Non-Work Related Issues, Journal of Managerial Issues.
- [28] Ongori, H. (2007); A Review of the Literature on Employee Attrition, African Journal of Business Management pp. 049-054, June 2007.
- [29] Rahman, A., Vaqvi Raza, S.M.M. and Ramay Ismail, M. (2008), Measuring Attrition Intention: A Study of IT Professionals in Pakistan, International Review of Business Research Papers, Vol. 4 No.3 June 2008 pp.45-55.
- [30] Siong Z.M.B, et al (2006); Predicting Intention to Quit in the Call Center Industry: Does the Retail Model Fit, Journal of Managerial Psychology, Vol 21, No 3, pp 231-243.
- [31] Steijn, B. and Voet, J (2009); Supervisors in the Dutch Public Sector and their Impact on Employees, EGPA Annual Conference, Malta, September 2-5 2009.
- [32] Tan, J., Tan, V and Tiong, T.N. (2006); Work Attitude, Loyalty, and Employee Attrition, Singapore Institute of Management, National University of Singapore.
- [33] Zhou, H., Long Lirong, R. and Wang Yuqing, Q. (2009); What is the Most Important Predictor of Employees' Attrition Intention in Chinese Call Centre: Job Satisfaction, Organizational Commitment or Career Commitment?, International Journal of Services Technology and Management, Vol 12, No 2, 2009, pp 129-145.

