Employee Turnover Analysis with Application of Data Mining Methods

K. Tamizharasi[1], Dr. UmaRani[2]

[1]Research Scholar,
Periyar University,
[2]Associate Professor,
Sri Saradha College for Women, Salem

Abstract- **Employee turnover is a usual thing in any business activities. They turn over for various reasons. But the companies have to monitor this issue because while hiring new employees there is training cost and associated hiring may increase. The rising cost of employee turnover confirms the need to provide more resources to finding solutions to the problem. Existing research confirms that the costs associated with losing a good employee and training a new one can equal 1.5 times the salary of the exiting employee. The study utilized data extracted from current employees by questionnaires and data of exiting employees of the Company, which included the individual reasons given for leaving the organization.To understand better the problem of turn over, the data are processed by the application of data mining methods: logistic regression, decision trees and neural networks. The models are built according to the SEMMA methodology and then compared to select the one which best predicts the employee turn over.**

Keywords:- Data Mining, Stress Management, Business Process Outsourcing, Turnover Analysis, Most Frequent Value

I. INTRODUCTION

Today's economic transformation, new information and communications technologies, and the globalization of the economy and labour market have led to labour market segmentation (J. Guichard, M. Huteau, 2005). According to the theory of segmentation, there is not one, but many isolated labour markets (Tanguy, 1986). It is often in search of work and in order to meet not only the economic needs that potential employees define their goals and aspirations outside their own local environment. Awareness of the existence of a foreign labour market, particularly strongly recognized in our region, strengthened by interconnections and the migration network, can make young people arrange their personal and professional lives taking into account such a perspective, especially as their parents, relatives and often also they themselves have already experienced migration (Cf. A. Krasnodębska, 2008).

There is a tendency that employees who have been sent to work abroad are more likely to seek for new job opportunities than the ones who have not (Stroh 1995). Black and Gregersen (1999) show in their study that 25% of the repatriates left their companies   within one year of repatriation which is twice as much as the once who have not experienced expatriation. There are different reasons for the resignation of the employees. Some suggest that the repatriates do not see any career opportunities in the companies they are working for (Paik et al 2002), others argue that the main reason is the lack of a repatriation program (Hurn 1999). Last but not least, during the employees' stay abroad there have been many changes in the home countries and also in the company. They have not been kept up-to-date with these changes and moreover, the employees have changed during their stay. Therefore, the expatriates have certain expectations and if these expectations are not met there is a possibility of resignation (Stroh et al. 1998).

Employees resign from organizations for a variety of reasons: compensation, future career opportunities, retirement, working conditions, quality of manager, job stress and impact on health, number of working hours, job security, and lack of a collaborative work environment (Carey & Ogden, 2004; Westcott, 2006). The focus of these studies was on reasons for employee attrition from companies in general, not during times of change or mergers. Turnover has been associated with poor communication both during mergers and separate from them (Carey & Ogden, 2004; Thornton, 2001). According to one study, "companies that communicate most effectively are more likely (51.6%) toreport turnover rates below those of their industry peers (33.3%)" ("Goodcommunication", 2004). Although inadequate communication is not cited as a directreason for departures in past studies, it may be related to attrition.

Employee turnover is a part of normal business activity; employees come and go as their life situations change. Employers realize this and, indeed, firms typically have entire departments devoted to the management of human resources in order to make the transition as painless as possible for both management and employee and to minimize the associated hiring and training costs.

Non job-related causes of employee turnover are generally out of the employer's control. Non job-related causes of employee turnover are those things in the employee's personal life that impact their performance in the workplace. Examples of these would be relocation, family problems and chemical abuse. Although these causes are not directly within the employer's control, some organizations

have sponsored responsive programs for the non job-related category such as employee assistance programs and stress management training that better prepared employees to deal with personal issues that impact their work performance.

Largest percentage (67.57%) of respondents left due to job related reasons, followed by 27.03% who showed personal reasons namely family, health, further studies reason etc. The job related reasons included: dissatisfaction with retail, dissatisfaction with salary, better employment non-retail, no career growth and scheduling problems.

A gender based analysis revealed that, males are more prone to leaving a job than that of the females. Roughly 74% of the males left the organization whereas the number of females was 26.0%. This reflects the overall gender imbalance evidenced in organization.

Previous research findings indicated that some causes of employee turnover are job-related factors that are somewhat within the direct control of the employer. Examples of such factors would be dissatisfaction with working conditions, supervising conflicts, scheduling conflicts or salary discrepancies. Understanding the causes of job-related turnover is crucial in being able to identify problems within an organization that might be controlled by the employer. Corrective steps taken in this area included training programs for supervisors, clarification of the employee's purpose or role and identifying scheduling solutions (Ulschak&Snowantle, 1992).

The paper is structured as follows: research settings are presented in the second section with data extraction and transformation procedures and research methodology given as separate subsections; third section states research findings while fourth section concludes the paper.

II. RESEARCH SETTINGS

Modern information systems collect daily large quantities of various data from different domains and sources of various forms and contents. In practice, there is a need for methods, techniques, and tools that can search vast data quantities, recognize patterns and present them on the level of concrete reports. In such complex requirements, the classic analytical approach is not sufficient as it is difficult to set a general mathematical model. A wide range of tools for collecting, storing, analyzing and visualizing data is defined as business intelligence (Michalewicz et al 2007).

To comprehend better the employee turnover, statistical data processing will be performed and some data mining methods will be applied. In the first segment, graphs will be used to present the basic information on the structure of students and obtain directions for a detailed analysis of dropouts. In the second segment, the analysis will be carried out by use of logistic regression, decision trees, and neural networks. Models will be built according to the SEMMA methodology and compared to select the one which best predicts the student dropout.

For pre-processing, the SQL language is used to perform a query over Sybase database, while particular data grouping

are carried out in Microsoft Office Excel 2003. SPSS 13.0 is used for designing clustered bar graphs, while data mining is conducted in SAS 9.1 Enterprise Miner. The software choice is SAS, which in the area of business intelligence dominates in advanced analytical solutions.

Data Extraction and Transformation

Transaction data on employee are collected through the various corporate companies employees real dataset. The data are stored in the Sybase database. There are two databases: the central database and the replicated database for the web. Updating of the replicated database is carried out when required, while the essential data, such as marks, are updated immediately. The system is based on the desktop and web components. The desktop section integrates all functions, while the web section is used within the system. Currently, there is no data warehouse, but it will be introduced in the near future to improve the reporting system and allow creation of ad-hoc reports. For web reporting, the replicated database is used while other reports are created from the central database. SQL and procedural programming are used for reporting at the moment.

To perform the analysis, the crucial tables in the database are EMPLOYEE– the basic data on the employees. For this analysis, the following attributes are separated from the database: RespondentID, CollectorID, Start Date, End Date,Name, Designation, Age, and Company. Besides these data collected at the enrolment, the analysis also includes attributes referring to the questioner process. The analysis is carried out on the sample of 2000 employees.

III. RESEARCH METHODOLOGY

A number of analytical methods are used for data mining. The standard types include regression (normal regression for prediction, logistic regression for classification), neural networks and decision trees (Olson and Delen, 2008). Different data mining methodologies show that the set of activities performed by the analyst can be presented as a series of logical steps or tasks. SEMMA (Sample, Explore, Modify, Model, and Assess) was developed by the SAS Institute which is also the producer of the data mining platform that uses the same methodology - SAS Enterprise Miner (SAS Institute, 2004; Matignon, 2007).

The acronym SEMMA signifies: Sampling, Exploring, Modifying, Modelling and Assessment. Starting with the statistically representative data sample, SEMMA allows the application of statistical and visualization techniques, selection and transformation of most significant predictor variables, modelling of variables to predict output, and eventually confirmation of model credibility.

To create classification models of neural networks, decision trees and logistic regression, it is necessary to follow the steps of SEMMA methodology. The model, developed in SAS Enterprise Miner, based on SEMMA methodology.
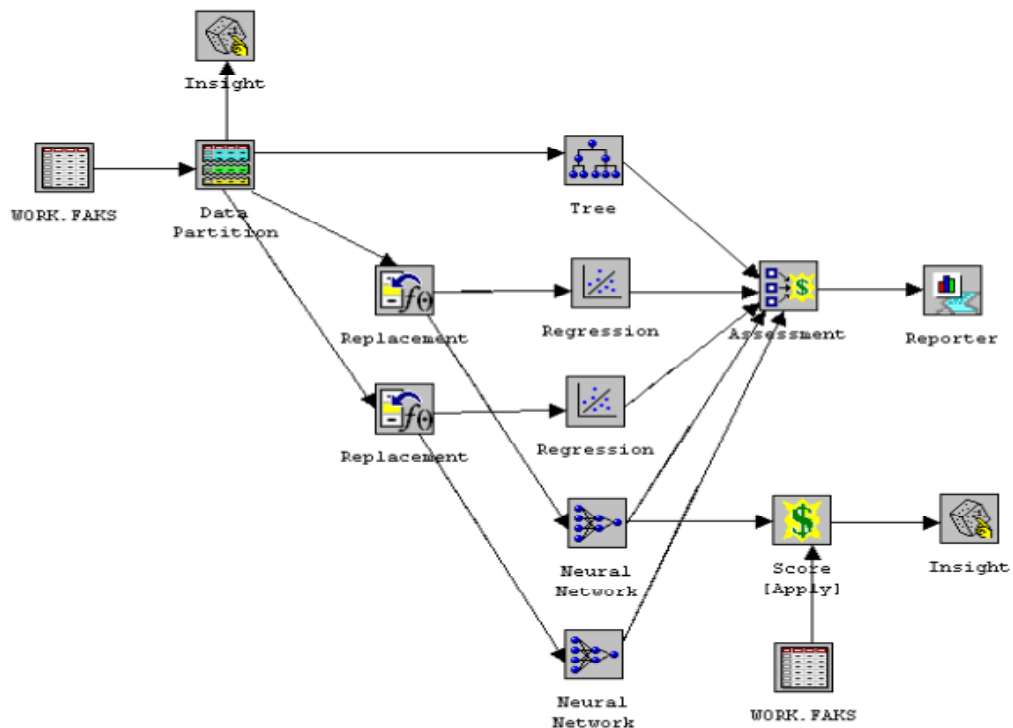
Figure 1: Model developed in SAS Enterprise miner based on SEMMA methodology

First, the input data are imported from the excel database and selected attributes that are analyzed in more detail. The turnout attribute is marked as the target variable that can assume two conditions: 1 for the employees who turnover and 2 for the employee who do not. The selected input variables are all those that are mentioned in the section on data extraction. In the Sample step, besides defining the role of variables in the model, it is also necessary to define one of the measuring scales (nominal, ordinal, and interval). This step also allows the testing of distribution for each variable.

Enterprise Miner in the Data partition node samples the input data and distributes them into training, validation and testing data (SAS Institute, 2003). The chosen method is Simple Random, in which each observation has the same possibility to be sampled.

In the step Explore, the node Insight is included. According to the settings for this analysis node, the sample size is 2000, and if it is less than that as in this case, the entire set of data is included in the analysis. In this example, 40% of the data is taken for training from the entire data set. This node offers numerous possibilities of data visualization. From the step Modify, according to the SEMMA methodology, two Replacement nodes are inserted to explore the effect of various data replacement methods on the logistic regression and neural networks models. In the upper node Replacement in Figure 1, the missing values replacement method is selected by Mode, or the most frequent values, while in the lower node, the selected method is Tree replacement which adjusts replacement values by using the decision tree.

Decision trees are a highly flexible modelling technique. For instance, to build regression models and neural networks models, the missing values have to be inserted into training data while decision trees can be built even with missing values. Decision trees are intended for the classification of attributes regarding the given target variable (Panian and Klepac, 2003). Decision trees are attractive because they offer, in comparison to neural networks, data models in readable, comprehensible form – in fact, in the form of rules. They are used not only for classification but also for prediction (Gamberger and Šmuc, 2001).

Logistic regression is an analysis of asymmetric relations between two variable sets of which one has the predictor status and the other criterion status (Halmi, 2003). The dependent variable is dichotomous and marked by values 0 and 1, while the independent variables in logistic regression may be categorical or continuous (Hair, Anderson and Babin, 2009).

Neural networks behave very well in more complex classification problems. Their disadvantage, in comparison to simpler methods, is the relatively slow and demanding process of model "learning" (optimization of weight factors) (Gamberger and Šmuc, 2001). Neural networks are a powerful tool in trend prognostics and predictions based on historical data. In data mining, neural networks are often combined with other methods because if used alone, they can hardly guarantee a good interpretation of results (Panian and Klepac, 2003).
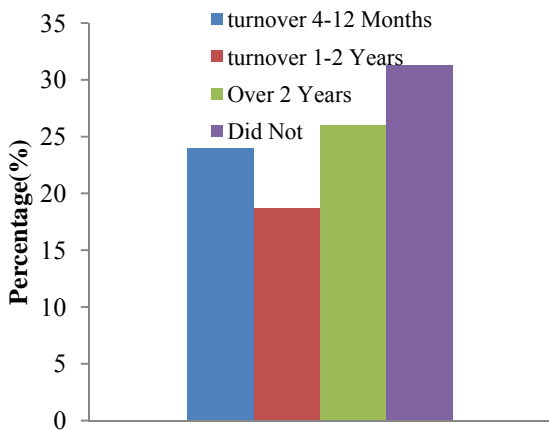
Figure 2: Percentage Distribution of Employees Turnover

The figure 2 shows the percentage of employees turned over to other jobs for various reasons in a certain period of 4 to 12 months, 1 to 2 year, over 2 years and still not. This gives a idea that many employees moves over in 4-12 months or else after 2 years more than the other.

One question on the interview questionnaire asked the respondents to indicate their reason for leaving the organization. When all the data was tabulated, different reasons for termination were identified. The reasons for termination were identified and then collapsed into few categories based on the similarities found. The individual and collapsed category results are given in figure 3.
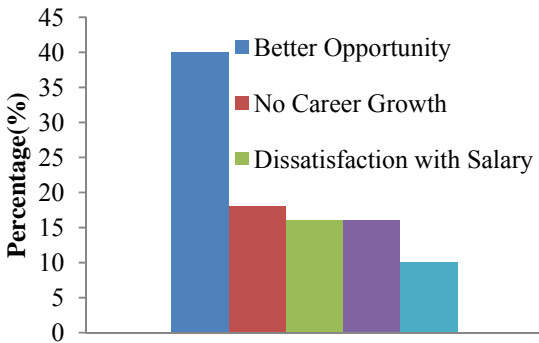
Figure 3: Reasons for Termination

From the figure 3, the major reasons for termination of the job is because of their better opportunity, then due to non career growth, then because of dissatisfaction on salary and work conditions and very few health issues.

IV. RESEARCH FINDINGS

Application of basic statistical methods is used to study the employee. It is found that women turnover comparatively less than men. To carry out a more detailed turnover analysis and separate the attributes which have the highest effect, decision trees are used.
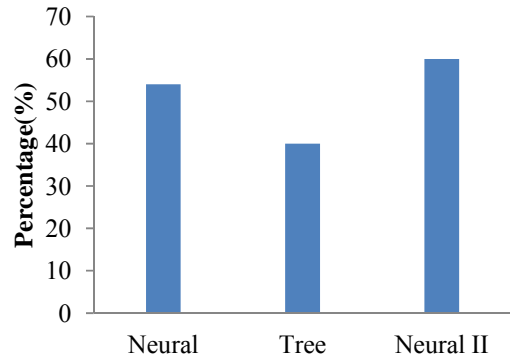
Figure 4 Evaluation and comparison of model I

The Figure 4 represents the basis or the 40% of employees from the original population that turnover. The graph shows that the model Neural-2 is evaluated as the best one in comparison to all the other models (it is positioned upwards). Other models are more difficult to interpret due to their overlapping. Figure 4 also shows that neural networks provide better results when using the missing values replacement method "Tree imputation" rather than "Most frequent value". The graph can be read in the following way: the plot Neural-2 shows that there are approximately 60% of employees that will turn over.

V. CONCLUSION

This paper has explored the application of the data mining methods in employee turnover. The problem of most employee turnover warrants consideration. The organization may contribute to the problem or chose to work to decrease employee turnover. By realizing the implications of actions involving employees and the organization can trustfully have a favorable impact on employee turnover. Organization must learn to adjust their traditional practices to avoid compounding the employee turnover problem.

Future research could examine reasons for termination of employee, as this would allow for any seasonal fluctuations in employment. Development of a more specific data mining tool which would address such factors as the existence of support systems among employees, changes in organizational strategies and inconsistencies in job expectations based on initial job which would provide valuable data for companies to retain their employees.

REFERENCES

[1] Black J &Gregersen B 1999, 'The right way to manage expats', Harward Business Review, vol. 77, no. 2, pp. 52-63.

[2] Carey, D. C., & Ogden, D. (2004). The human side of M&A: Leveraging the most important factor in deal making. New York: Oxford University Press.

[3] Cf. A. Krasnodębska, Occupational preferences of Opole students and the issue of labour-related migration abroad, WydawnictwoInstytutŚląski, Opole 2008.

[4] Gamberger, D., Šmuc, T. (2001): Poslužiteljzaanalizupodataka (http://dms.irb.hr). Zagreb, Hrvatska: InstitutRudjerBošković, Laboratorijzainformacijskesustave.

[5] Good communication can help boost both your bottom line and employee retention. (2004, September). Contractor's Business Management Report, 5-7.

[6] Hair, J., Anderson, R., Babin, B. (2009): Multivariate Data Analysis, Prentice Hall.

[7] Halmi A. (2003): Multivarijantnaanaliza u društvenimznanostima, Alinea, Zagreb

[8] Hurn, J 1999, 'Repatriation - the toughest assignment of all', Industrial and Commercial Training, vol. 31, no. 6, pp. 224-228.

[9] J. Guichard, M. Huteau, Counselling Psychology, OficynaWydawnicza "Impuls", Krakow 2005, p.14.

[10] Matignon, R. (2007): Data Mining Using SAS Enterprise Miner TM, John Wiley & Sons, Inc., Hoboken, New Jersey

[11] Matignon, R. (2007): Data Mining Using SAS Enterprise Miner TM, John Wiley & Sons, Inc., Hoboken, New Jersey

[12] Olson, D.L., Delen, D. (2008): Advanced Data Mining Techniques, Springer-Verlag, Berlin Heidelberg.

[13] Paik, Y, Segaud B & Malinowski C 2002, 'How to improve repatriation management', International Journal of Manpower, Vol. 23, no 7, pp. 635-675.

[14] Panian, Ž., Klepac, G. (2003): Poslovnainteligencija, Masmedia

[15] SAS Institute (2004): Getting Started with SAS Enterprise Miner 4.3, Second Edition, SAS Institute Inc., Cary

[16] Stroh, L 1995, 'Predicting turnover among repatriates: can organizations affect retention rates?', The International Journal of Human Resource Management, vol. 6, no.2, pp. 443-456.

[17] Stroh, L, Gregersen, B & Black, J 1998, 'Closing the Gap: Expectations Versus Reality Among Repatriates', Journal of World Business, vol. 33, no. 2, pp. 111-124

[18] Tanguy, 1986, see: ibidem, p. 21.

[19] Thornton, S. L. (2001, October/November). How communication can aid retention: Making communication a management priority. Strategic Communication Management, 24-27.

[20] Ulschak, F.L., &Snowantle, S.M. (1992). Managing employee turnover; A guide for health care executives. Chicago, Illinois: American Hospital Publishing.

[21] Westcott, S. (2006, April). Goodbye and good luck. Inc. Magazine, 40-42.