A Novel Approach to Control the Employee's Attrition Rate of an Organization

[1]P. Saranya Devi, [2]Dr. B. Umadevi
[1]Research Scholar, [2]Assistant Professor & Head
[1]P.G. & Research Department of Computer Science
[1]Raja Doraisingam Govt. Arts College, Sivaganga, Tamil Nadu, India
[2]P.G. & Research Department of Computer Science
[2]Raja Doraisingam Govt. Arts College, Sivaganga, Tamil Nadu, India

_____

Abstract

Every organization has its own productivity and strength which stands of the legs of the employees. Keeping regular employee is a great challenge for all organization in the competitive world. The controlling the attrition rate is a great challenge. An employee leaving any organization is carrying not only the tactic knowledge but the potential experience also. The competency level of the individual is makes the strength of the organization. These study interpreters the employees' attrition rate through the related attributes. The IBM Watson Analytics Community Human Resource (HR) data for employee attrition from UCI repository were collected for the analysis. A prediction based on classification algorithm is applied to predict the attrition rate of the employee.

Keywords: KDD, Attrition Rate, HR, Classifier.
_____

1. Introduction

The outcome of many research shows that the most valuable asset and important resource in organizations are their employees. Now a day due to increased competition and improved requirement in employees' proficiency determines the attrition rate. The employee attrition is considered to be a serious issue for organizations. The cost of searching and training employees is very high. Organizations need to search, hire and train new employees. Loss of experienced workers especially high performers is difficult to manage and is negatively related to the success and performance of organizations. The study focuses on the variables that may lead to control the attrition rate of the employee.
Identifying the most significant factors which affects employee attrition is vital for implementing business strategies by selecting and adjusting proper improvement activities for retaining and hiring new employees. It highly gives importance for predicting plausible employee attrition by using data mining algorithms.

The problem of employee turnover has turn to eminence in organizations because of its pessimistic impacts on issues on work place self-esteem and efficiency. [1] The organizations deal with this problem is by predicting the risk of attrition of employees using machine learning techniques thus giving organizations to take pro-active action for retention.

In this paper, the problem of employee [2] turnover and the key machine learning algorithms that have been used to solve it are discussed. The paper is structured as follows. The section one describes about introduction. Section 2 deals about background study and its related works. The methodology of the research work is explained in section three. In section four portraits the Experiment results. Finally the paper is concluded in last section.

2. Background and Related Works

Data mining is the most promising field for information and knowledge discovery. It started to be an interest goal for information diligence, due to huge data containing large amounts of unknown knowledge. With data mining techniques, such knowledge can be extracted and accessed transforming the databases tasks from storing and retrieval to learning and extracting knowledge.

Data miming has its own set of techniques that can be used to mine relevant and attractive knowledge from data. Data mining has several techniques [3] such as association rule mining, classification, prediction, and clustering. Classification techniques are supervised learning techniques that classify data item into predefined class label. It is one of the most useful techniques in data mining [4] to build classification models from an input data set. The classification techniques commonly build models that are used to predict future data trends.

2.1 Literature Review

In this paper, diversified approaches using various data mining techniques are collected to analyse the employee attrition rate at various levels. The study related to data mining for extracting and predicting the employee's attrition rate used in various models and the comprehensive literature review of various researchers' works are stated below.

Qasem A, A.Radaideh and Eman A Nagi et al, has applied data mining techniques to build a classification model to predict the performance of employees [5]. They adopted CRISP-DM data mining methodology [6] in their work. The Decision tree was the main data mining tool used to build the classification model, where several classification rules were generated. They validated the generated model; several experiments were conducted using real data collected from several companies. The model is intended to be used for predicting new applicants' performance.

Amir Mohammad Esmaieeli Sikaroudi, [7] RouzbehGhousi and Ali EsmaieeliSikaroudi et al, implemented knowledge discovery steps on real data of a manufacturing plant. They chew over many characteristics of employees such as age, technical skills and work experience. They used to find out importance of data features is measured by Pearson Chi-Square test.

John M. Kirimi and Christopher A. Moturi et al, [8] proposed a prediction model for employee performance forecasting that enables the human resource professionals to refocus on human capability criteria and thereby enhance the performance appraisal process of its human capital.

Rohit Punnoose and Pankaj Ajit et al, explored [9] the application of Extreme Gradient Boosting (XGBoost) technique which is more robust because of its regularization formulation. [10] Data from the HRIS of a global retailer is used to compare XGBoost against six historically used supervised classifiers and demonstrate its significantly higher accuracy for predicting employee turnover.

Hamidah Jantan, [11] Abdul Razak Hamdan, and Zulaiha Ali Othman et al, suggested the potential HR system architecture for talent forecasting by using past experience knowledge known as Knowledge Discovery in Database (KDD) or Data Mining.

3. Methodology

The applications of various data mining techniques which is adopted as a methodology to predict employee attrition. The most predictive data models applied are artificial neural networks, Naïve Bayes, K-Nearest Neighbour (KNN), Support Vector Machine, logistic regression, classification trees, classification and regression trees and discriminant analysis. The main objectives of the studies are:

(i) To evaluate the effectiveness of different models with respect to their predictive accuracy.
(ii) To identify the factors that influence employee attrition.
(iii) To develop a predictive models for employee attrition.

The following section describes the popular models which are used to predict the employee attrition. The steps involved in Predicting the Employee attrition rate is explained in Figure 1.

3.1 Naïve Bayes

The Naïve Bayes is a popular classification technique that it attracts for its simplicity and performance. Naïve Bayes performs classification based on probabilities arrived, with a base assumption that all variables are conditionally independent of each other. To estimate the parameters (means and variances of the variables) necessary for classification, the classifier requires only a small amount of training data. It also handles real and discrete data.

The underlying logic behind Bayes' rule for machine learning is as follows: To train a target function f: X → Y, which is the same as, P (Y|X), we use the training data to learn estimates of P (X|Y) and P(Y). Using these estimated probability distributions and Bayes' rule new X samples could then be classified.

3.2 J48 (C 4.5)

The J48 (C4.5) technique is one of the decision tree families that can produce both decision tree and rule-sets and construct a tree for the purpose of improving prediction accuracy. The J48 classifier is among the most popular and powerful decision tree classifiers. The J48 creates an initial tree using the divide-and-conquer algorithm.

3.3 Random Tree

Random Tree algorithm is a popular tree based ensemble learning technique. The bagging type used here is "ensembling". In bagging, successive trees do not depend on earlier trees each is independently [12] constructed using a different bootstrap sample of the data set. In the end, a simple majority vote is taken for prediction. Random Trees are different from standard trees in that for the latter each node is split using the best split among all variables. In a random tree [13], each node is split using the best among a subset of predictors randomly chosen at that node. This additional layer of randomness makes it robust against over-fitting.
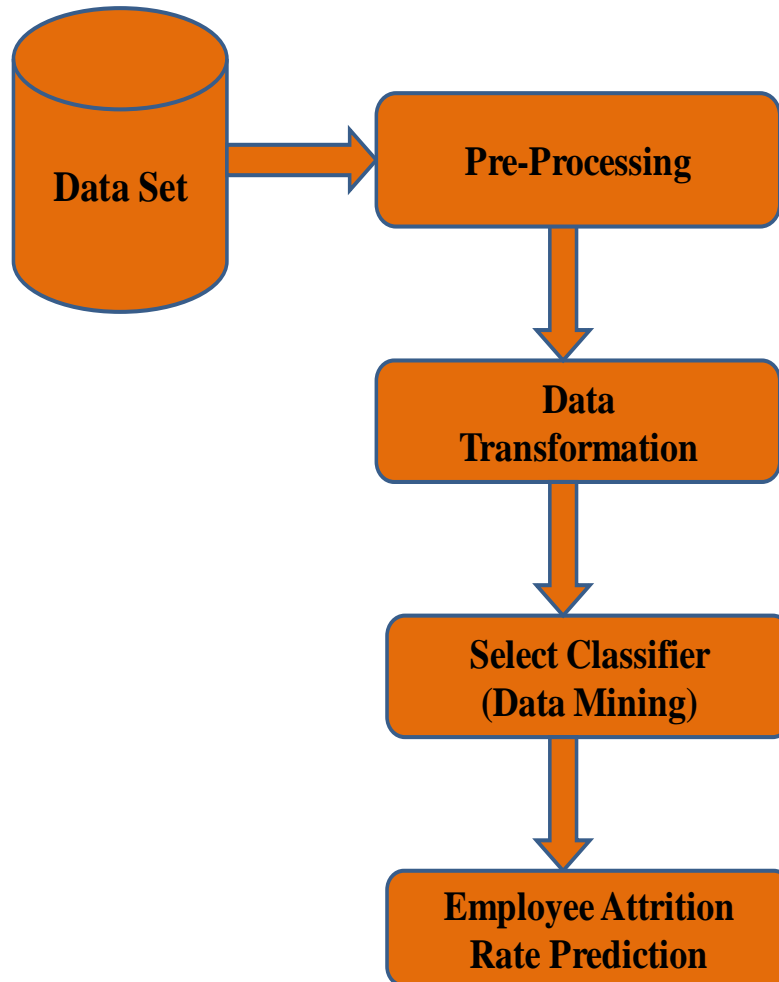
Figure 1: Steps involved in Predicting the Employee Attrition Rate

4. Experiment Results

The data set is collected from UCI for Human Resource (HR) Community IBM Watson Analytics data. From the data set 300 samples are took for analysis for the attributes included mentioned below. The attributes are Age, Department, Education, Gender, Job-Role, Job-Satisfaction, Monthly Income, Performance Rating, Relationship-Satisfaction, Years-At-Company, Years-Since-Last-Promotion, Years-With-CurrManager and Attrition.

The data preprocessing is applied to extract the best attributes through cfs-subset evaluator and best-first search algorithm. The three classification algorithm such as Naïve bayes, J48 (C4.5) and Random Tree are applied on the data set. The objective is for predicting the employee attrition rate in an organization. The status of the attrition rate is either "Yes" or "No" by the different classifier is given in Figure 2 to Figure 4.
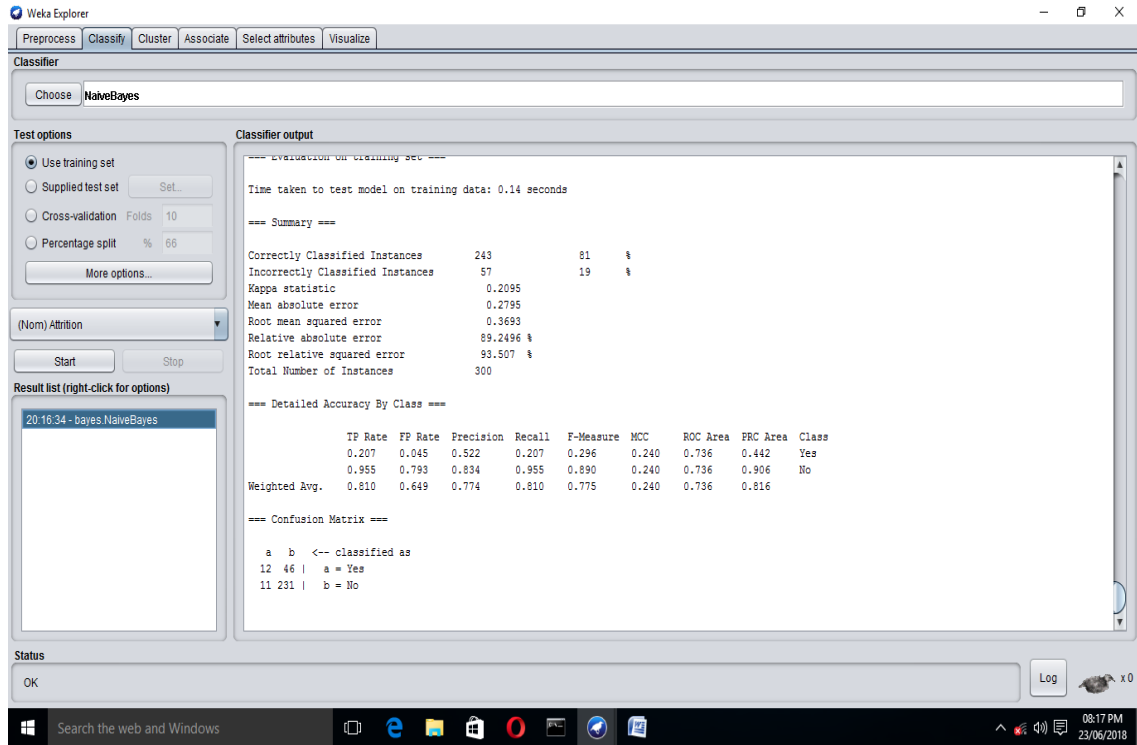
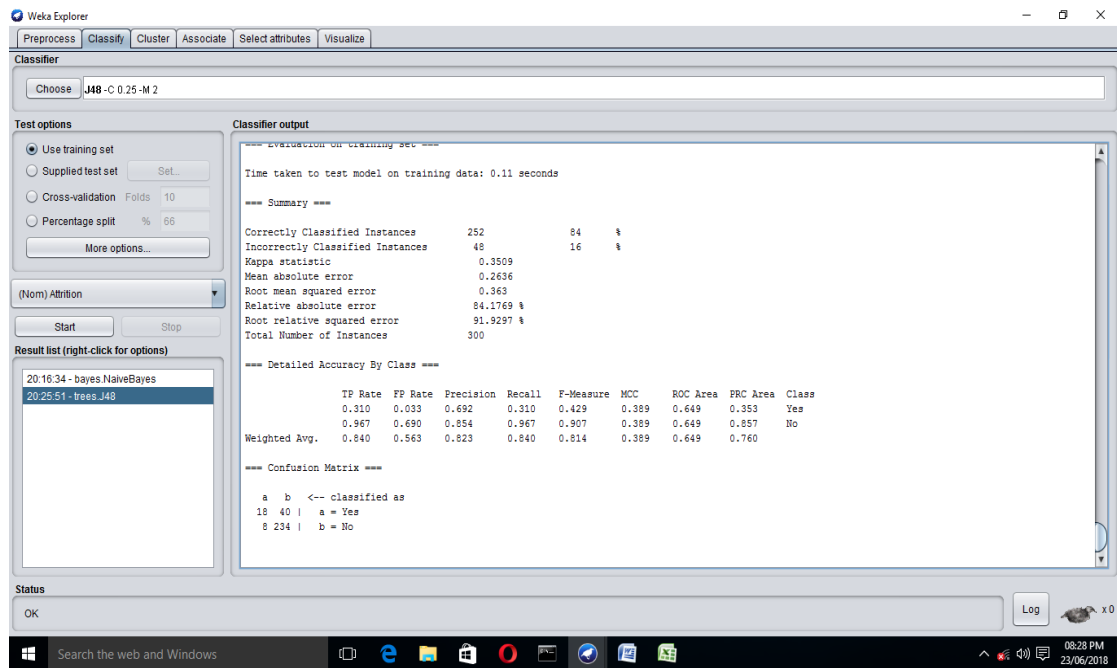**Figure 2 :** Attrition Rate Classification by Naïve Bayes

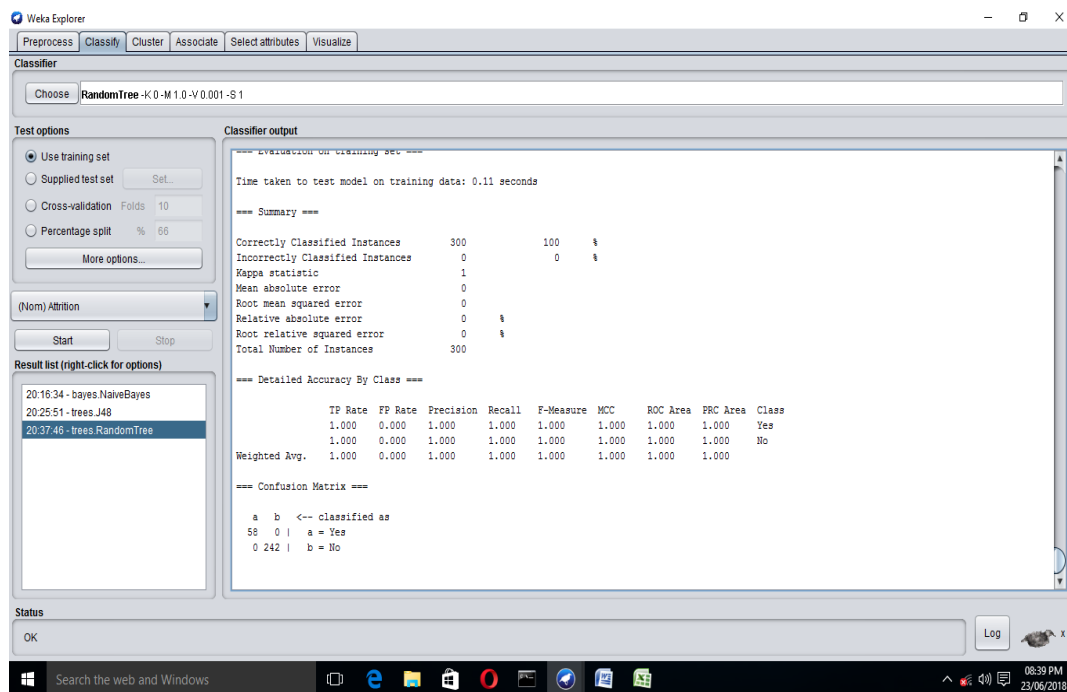**Figure 3 :** Attrition Rate Classification by J48 (C 4.5)

Figure 4 : Attrition Rate Classification by Random Tree

The algorithmic comparison by Naïve bayes, J48 (C4.5) and Random forest interprets the accuracy of the attrition rate of being inside as well as employee leaving the organization by the accuracy measurement parameters. The comparisons of Attrition rate of different classifiers are given in Table 1 and Table 2 and also the graph is shown Figure 5 and Figure 6.

Table 1 : Comparison of Classifiers for Attrition Rate – "Yes"

Classifier	True Positive (TP)
Naive-Bayes	0.207
J48 (C4.5)	0.310
RandomTree	1.000

Table 2 : Comparison of classifiers for Attrition rate – "No"

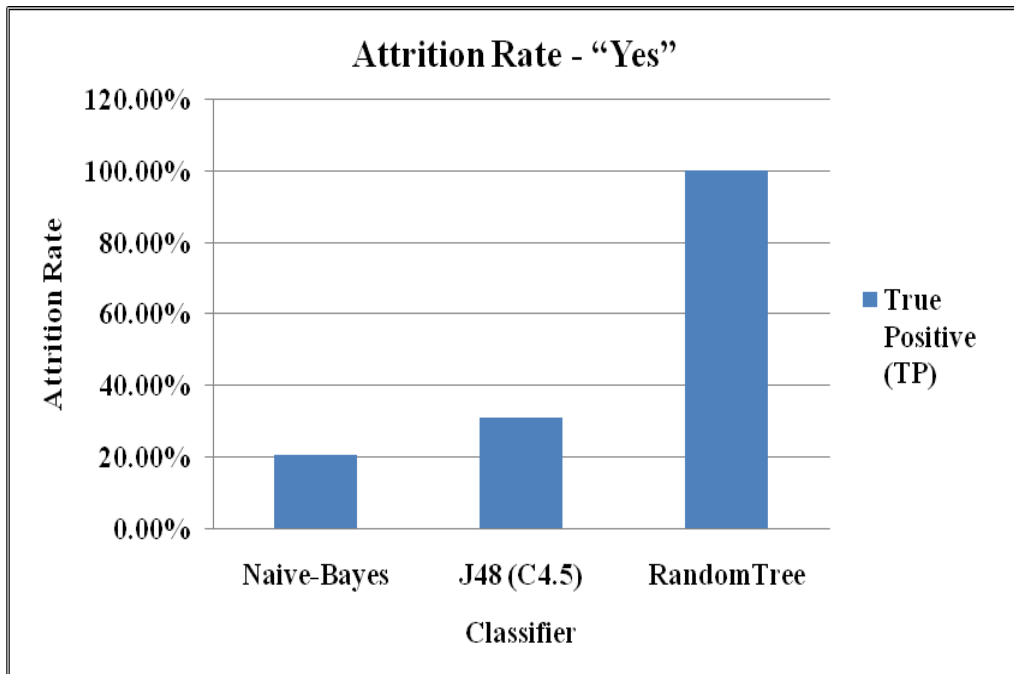| Classifier	False Positive (FP)
Naive-Bayes	0.793
J48 (C4.5)	0.690
RandomTree	0.000

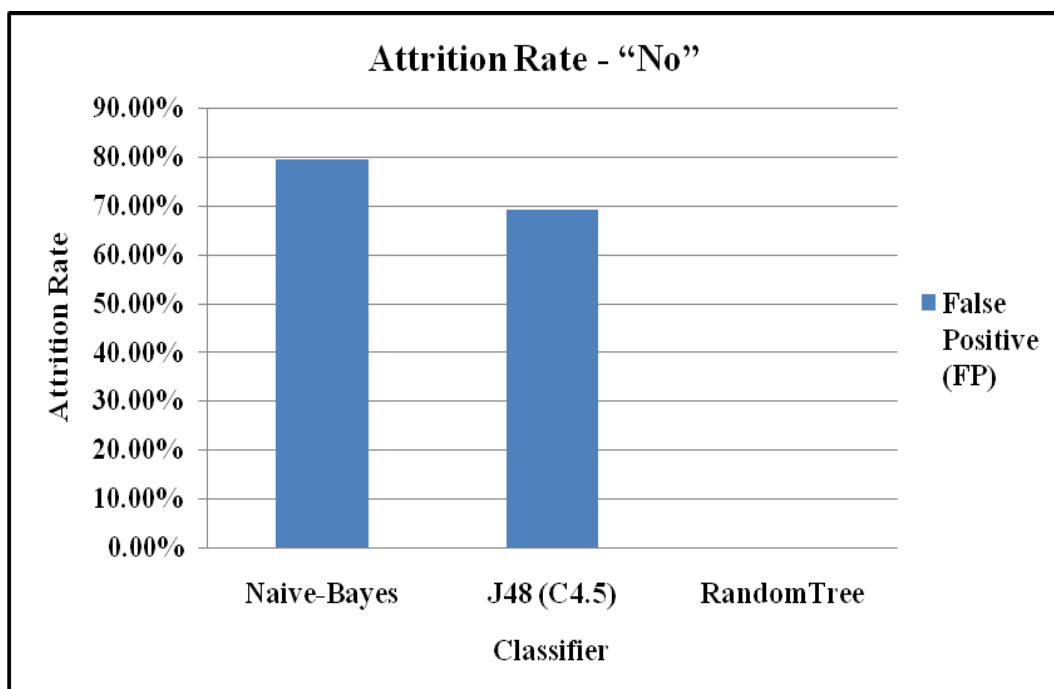Figure 5 : Attrition Rate – "Yes" by Classifiers

Figure 6 : Attrition Rate – "No" by Classifiers

5. Conclusion

Human Resource is the main pillar for any organization. The growth level as well as market penetration are duly depends on the strength of the employees. Now a days due to increased population and people with high competency makes great success for any firm. But the prime issues which are normally addressed in any organization are only the attrition. This is a great challenge as well as retention is also the prime task. In accordance with the attrition rate we took data from UCI for HR Community IBM Watson Analytics data. We took three mining algorithms like Naïve bayes, J48 (C4.5) and Random Tree respectively. From the analysis the Employees attrition rate is analyzed classified by the two parameters such as "Yes" and "No". Yes interprets people leaving the organization and NO says staying with the same organization. Among three algorithms the percentage of employees leaving from the firms is 20.70% by Naïve bayes, 31% by J48 and 100% by Random Tree. The retention is 79.3% by Naïve Bayes, 69% by J48 and 0% by Random tree. Among the analysis the Naïve bayes gives better performance rather than other two algorithms. Now the organizations can make right decision with the results interpreted by Naive Bayes.

References

[1] Alao D. and Adeyemo A. B, 17th March 2013, "Analyzing Employee Attrition Using Decision Tree Algorithms", Computing, *Information Systems and Development Informatics* , Volume 4 No. 1.

[2] Amir Mohammad Esmaieeli Sikaroudi, Rouzbehghousi and Ali Esmaieelisikaroudi, Autumn 2015, "A Data Mining Approach To Employee Turnover Prediction" (Case Study: Arak Automotive Parts Manufacturing) , *Journal Of Industrial And Systems Engineering* , Volume. 8, No. 4, Pp106-121.

[3] Ashish Kumar Dogra and Tanuj Wala, , June 2015 "A Comparative Study Of Selected Classification Algorithms Of Data Mining", *International Journal Of Computer Science And Mobile Computing* (IJCSMC), Volume. 4, Issue. 6, Pg.220 – 229.

[4] Brijesh Kumar Bhardwaj and Saurabh Pal, April 2011 "Data Mining: A Prediction For Performance Improvement Using Classification", *International Journal Of Computer Science And Information Security* (IJCSIS), Volume. 9, No. 4.

[5] Renuka Agrawal, Jyoti Singh, and Zadgoankar .S, "Formative Assessment For Performance Evaluation Of Faculty Using Data Mining", *International Journal Of Advances In Electronics And Computer Science*, ISSN: 2393-2835.

[6] Hossein Alizadeh, Buinzahra Branch and Islamic, 2016 ,"Introducing A Hybrid Data Mining Model To Evaluate Customer Loyalty", *Engineering, Technology & Applied Science Research Volume. 6, No. 6, 1235-1240.*

[7] Amir Mohammad Esmaieeli Sikaroudi , Rouzbehghousi and Ali Esmaieelisikaroudi, 2015 "A Data Mining Approach To Employee Turnover Prediction" (Case Study: Arak Automotive Parts Manufacturing), *Journal Of Industrial And Systems Engineering* Volume. 8, No. 4.

[8] Anjali A. Dudhe and Sachin Sakhare .R, January 2018, "Teacher Ranking System To Rank Of Teacher As Per Specific Domain" , *Journal On Soft Computing ICTACT*, Volume: 08, Issue: 02, Issn: 2229-6956.

[9] Rohit Punnoose and Pankaj Ajit, 2016 "Prediction Of Employee Turnover In Organizations Using Machine Learning Algorithms", *International Journal Of Advanced Research In Artificial Intelligence,* (IJARAI) Volume. 5, No. 9,.

[10] Hamidah Jantan, Abdul Razak Hamdan, And Zulaiha Ali Othman, 2009, "Knowledge Discovery Techniques For Talent Forecasting In Human Resource Application", *International Journal Of Industrial And Manufacturing Engineering* Volume:3, No:2.

[11] Rupesh Khare , Dimple Kaloya , Chandan Kumar Choudhary and Gauri Gupta , "Employee Attrition Risk Assessment Using Logistic Regression Analysis", International Conference On Advanced Data Analysis (IIMA), Business Analytics And Intelligence, Indian Institute Of Management.

[12] Krishnan Ramanathan, Anil Jadhav . Dr., 2016, "Intelligent Decision Support Systems - A Tool For Human Resource Allocation In Information Technology Projects", *International Journal Of Computer Science And Information Technologies*, (IJCSIT) Volume. 7 (5) , 2356-2364.

[13] Cheolyong Park, "Simple Hypotheses Testing For The Number Of Trees In A Random Forest" , 17 March 2010, *Journal Of The Korean Data & Information Science Society* 21(2), 371–377.

[14] Saranya Devi .P.and Umadevi .Dr. B.**,** April 2018.,"An Overview Of Organizational Growth By Human Resource Based On Knowledge Management", *International Journal of Creative Research Thoughts (IJCRT)*, Volume 6, Issue 2.

[15] Umadevi, B., Sundar, D and Alli, Dr.P. 10th June 2014, "Novel Framework For The   Portfolio Determination Using PSO Adopted Clustering Technique". *Journal of Theoretical and Applied Information Technology*. Volume. 64- No.

[16] Umadevi, B. Sundar, D and Alli, Dr.P. 26-28 Dec. 2013. "An Optimized Approach to Predict the Stock Market Behaviour and Investment Decision Making using Benchmark Algorithms for Naive Investors",*Computational Intelligence and Computing Research* (*ICCIC*). *IEEE International Conference.* Page(s):1 – 5978-1- 4799-1594-1.

[17] Umadevi, Dr. B and Dhanalakshmi, R. April 2017. "A Comprehensive Survey of Students Performance Using Various Data Mining Techniques". International Journal of Science and Research (IJSR) ISSN (Online): 2319-7064. Volume 6 Issue 4.

Authors' Biography

P. SARANYA DEVI is an M.Phil Research Scholar in PG & Research Department Of Computer Science, Raja Doraisingam Government Arts College, Sivaganga, Tamilnadu, India. Her research concentration include in Data mining, Machine Learning and its applications. She published paper in internatinal journal.

Dr. B. UMADEVI has received her Doctoral degree in Computer Science from Manonmaniam Sundaranar University, Tirunelveli, India. Currently working as Assistant Professor & Head- P.G and Research Department of Computer Science, Raja Doraisingam Government Arts College, Sivagangai-Tamilnadu, India. She has over 23 years of Teaching Experience and published her research papers in various International, National Journals and Conferences. Her research interests include Data Mining, Soft Computing and Evolutionary Computing. She got the Best Paper Award for her publication in the IEEE International Conference on Computational Intelligence and Computing Research held on 27th Dec 2013 at VICKRAM College of Engineering and Technology.