

Survey on Predicting Performance of An Employee using Data Mining Techniques

S. E. Viswapriya

Department of Computer Science and Engineering,
SCSVMV University, Kanchipuram,
Tamilnadu, India

Abstract:- Predicting analytics is an upcoming trend in human resources. Predictive data analytics are every where. It is in its essence a technology that learns from existing data and it uses this to forecast individual behaviour. Data mining plays an important role in the field of predictive analytics. A common and rather simple method to create a predictive model is decision tree. Human resource posse's large quantities of data. Analytics can mine data on candidate's personality, behavioural traits and skills to throw useful insights into whether he or she would be the right fit for the organization. This paper discuss about a brief literature survey on several papers published to predict employee performance using data mining techniques.

General Terms:- Data mining, decision tree, classification and clustering techniques.

INTRODUCTION

Data Mining is a set of method that applies to large and complex databases. This is to eliminate the randomness and discover the hidden pattern. We use data mining tools, methodologies, and theories for revealing patterns in data. There are too many driving forces present. And, this is the reason why data mining has become such an important area of study. We use data mining to automate the process of finding predictive information in largedatabases. Also, to identify previously hidden patterns in one step. Several types of data such as Relational database, data ware house, Advanced DB and information repository, object oriented and object relational databases, transactional and spatial databases, heterogeneous and legacy database, multimedia and streaming data base, text database were all used for mining.

DATA MINING IMPLEMENTATION PROCESS

There are five implementation process in data mining. They are as follows:

1. Business Understanding – Which is used to establish business and data mining goals.
2. Data Understanding – Data can checked to find whether it is appropriate for data mining goals.
3. Data Preparation – Data can be prepared well by sequential process such as selecting, cleaning, transforming, formatting, anonymizing and constructing.
4. Data Transformation – Data transformation can be done by smoothing, aggregation, generalization, normalization and attribute construction.
5. Modelling – Mathematical models are used to pattern the data.

DATA MINING TECHNIQUES

Classification:

The information about data and metadata has been analysed to retrieve relevant information. The main part of classification method is to classify the data in to different classes.

Clustering:

This method is used to find similarities and differences among the given data.

Regression:

This method is used to identify and analyse relationship between variables and also used to find the similarities on variables.

Association Rule:

The hidden part in data set and their associated data can be found by using this method.

Prediction:

By using past events or instances the future event will be predicted. The prediction can be done easily by combining all other data mining methods.

Data mining tools:

The two popular tools for data mining are as follows:

R Language:

This tool used for statistical, classical, time series analysis, Classification and graphical techniques.

Oracle Data Mining:

This tool widely used in generating detailed insights and to make predictions easily.

Data mining Applications:

Data mining has been mainly applied in the area of Communications, insurance, education, banking, manufacturing, retail, service providers, Ecommerce, Super market, Crime investigations, Bio informatics.

Advantages of data mining:

In order to get knowledge-based information, profitable adjustments made in companies, helps in decision making, easy to analyse huge amount of data in less time. It is cost effective and more efficient in finding the hidden patterns.

Disadvantages of data mining:

Many data mining software were difficult to work and needs more advance training to work on. There are several variant data mining tools are available with different algorithms. It is difficult to select the correct tool in correct situation. The result of any project by using data mining tools are not so accurate in some situations and so it may leads to serious consequences in certain conditions.

Artificial Neural Networks

We use data mining in non-linear predictive models. As this learn through training and resemble biological neural networks in structure.

Decision Trees

As we use tree-shaped structures to represent sets of decisions. Also, these rules are generated for the classification of a dataset. These decisions generate rules for the classification of a dataset. As there are specific decision tree methods that include Classification and Regression Trees and Chi-Square Automatic Interaction Detection (CHAID).

Genetic Algorithms

There are the present genetic combination, mutation, and natural selection for optimization techniques. That is design based on the concepts of evolution.

Nearest Neighbour Method. A technique that classifies each record in a dataset based on a combination of the classes of the k records. It in a historical dataset (where $k \geq 1$). Sometimes called the k-nearest neighbour technique. Rule Induction The extraction of useful if-then rules from data based on statistical significance.

ID3 Algorithm:

The entire data set has been used to create tree. This algorithm will build short and fast tree.

K Nearest Neighbour:

It is easy to implement for simple technique and suitable for multi model classes.

Naïve Bayes:

It is very simple and quicker than other models. It will not need more training on data set.

Support vector machine:

This algorithm deals with both linear and nonlinear data. The data is redundant if there is boundary in the given data set.

CART (Classification and regression techniques):

It is easy to handle numerical and categorical variables. It includes only significant variables.

2. COMPREHENSIVE REVIEW OF LITERATURE

A comprehensive literature review of various significant areas of predicting employee performance using data mining techniques has given in below tabular form (Table 1).

Table 1

Category	Year, Author(s)	Methodology	Key Findings
Survey of papers published in predicting employee performance to promote him	2010, Hamidahjantan, mazidahputeh,abdulrazak ha, zulaihaalimadan Othman	Classification Algorithms	Identified accuracy of classification technique in Data mining that helps to promote employee based on his performance.
	2017, jiechaocheng	CRISP-DM cycle process	Identified Data Mining Technique to enhance decision making and analyzing new patterns and relationships for organizations.
	2018, Rahul yedida, Rakshitvahiabhilash, rahulreddy, rahul j, deeptikulkarni	K- Nearest neighbors algorithm. It also includes Artificial neural network, Decision Tree, Logistic regression	Identified several methods of classification determined accurate results among different algorithms K – Nearest neighbor gives accurate results.
	2019, KedirEyasn, AbdulKadhir, FlueaAmena, Toltsa	Data Mining Classification Algorithms, CRISP-DM techniques, Hybrid Data Mining process Model.	Efficiency and Effectiveness of employees is determined using Data Mining Classification Technique with basic parameters. It does not include confidential information about employee.
	2018, Ananyasarker, SM. Shamim, Dr. M.D. Shahiduz Zama and Md. Mustafizur Rahman	K- means Clustering, Decision Tree Algorithms.	Identified inefficient employee, Magnitude of inefficiency and aids to eliminate inefficiency with a relatively easy to employee framework k - means clustering for partitioning the employee and Decision Tree Algorithm for classify employee and take appropriate decision quickly.
Survey of papers published in predicting Employee turnover	2019, Yuezaho, Maciej. Hryniewrcki, Francesca cheng, Boyang Fu, and Xiaoyuzhu.	Supervised Machine Learning method, Decision Tree, Random Forest, gradient Boosting trees, Logistic Regression, Support vector machine, Neural network, LDA, Naïve Bayes, K- Nearest neighbor.	Measured the effectiveness of various supervised machine learning algorithms. Described the evaluation criteria, Algorithm effectiveness and procedures that used in conducting the numerical experiments performed in this research. Discussed five evaluation metrics in the evaluation of supervised machine learning algorithms.
	2017, SubhamTupe, Chetan Mahajan, Dnyanshwaruplenchwar, Pratik Deo.	Entropy, Decision Tree Algorithm, ID3 Algorithm.	Entropy has been calculated and information gain. Decision tree is created by ID3 Algorithm. Using Data Mining as a tool we can handle data in supervised way.
	2019, Farhad Sheybani	CHAID Decision Tree algorithm	Identified disinterest of the employee to continue their work in the same organization due to lack of individuals pride. Determined the rate of dissatisfaction of job.
	2017, Nor Azziaty, Abdul Rahman, Kian lamtan, and chenkimlim	Supervised and unsupervised machine learning Algorithm such as K- Nearest Neighbor, Naïve Bayes, Decision tree, Neural network, Logistic Regression, Support vector machine, CRISP-DM model	This article helps to identify the skilled, knowledgeable and fulfilled employee. Proposed suitable classification model for predicting and assessing attributes of employees dataset to meet the criteria of work demanded by the industry.
	2019, Subhamkarande, Ajay shelake, sivagami M, Sharon Sophia.	Classifier techniques such as Support vector machine, Multi layer perceptron, Logistic Regression, Voting classifier, Apache Cassandra	Built an ensemble learning model which is a combination of Support Vector Machine, Logistic Regression, Random Forest. Based on accuracy, this model will be able to predict the turnover of employees. Final classification is based on Weighted average.
	2019, Snikhan, Khera, Divya	Supervised machine learning algorithm, Support Vector Machine.	The Employee data can be tested for this accuracy by using supervised machine learning classification models and hence validated.
	2019, ZarminaJaffar, Dr. Waheed Noor, ZartashKanwal	Data Mining Techniques such as J48, Naïve Bayes and Logistic Regression.	Correlation based method is used to ensure that connection between variable and components in the testing can be estimated. Classification step through training and testing data. Association also be used to reveal all the relationships in a large database. The accurate results can be derived from J48 decision tree algorithm.
	2019, Xiang Gao, Juhao Wen, Cheng Zhang	Decision tree algorithm.	The extraction of subsample from original samples is done using Random Forest algorithm. It classifies decision trees and implements simple vote. The degree of decreasing accuracy of Random Forest prediction is calculated by adding noise to each feature.
	2016, SujeetNarendramishraDevR	Data Mining Techniques	The necessity of Data mining Techniques in order to predict analytics for human resource management is discussed.

aghavendralama 2013, Akhilesh k Sharma, Dr. KamalijitLakhtaria, Santosh Viswakarma	CRISP Model, Weka	Predicting the professional skill development program or newly hired employees based on their current training needs. Discussed about various applications of data mining on predicting the employee performance.
2015,Amirah Mohamed Shahir, Wahidah Husain, Nuraini Abdul Rashid.	Data Mining Techniques, Classification algorithms such as Decision tree, Artificial Neural network, Naïve Bayes, K- Nearest Neighbor and Support Vector Machine	Analysed difference prediction methods for predicting students performance. Identified important attributes in predicting students performance while predicting students performance. CGPA plays an important attribute among other attributes list.
2014, Saurabh pal, Ajay Kumar pal	ID3 Algorithm, CART, LAD	Teachers performance has been evaluated. Identified two step model of classification techniques in Data Mining. Determination of class label attributes with the help of tuple. Identified the characteristics of dropout students by using Naïve Bayes Classification. Preprocessing techniques has been used to remove noise in data sets with the help of ID3 Algorithm. Data Preparation, Data selection and transformation are process that could be done data mining techniques.
2012, Stefan StrohmejerFrancal Piazza	Data Mining methods and Technology	Identified Domain Driven Framework contributions leads to variety of HR Domain problems and Data Mining methods.
2010,EWT.Ngai, Yong Hu, Y.H. Wong, Yijunchen, Xinsun	Classification Algorithm, Regression, Clustering, Prediction	Identified the detection of financial fraud. To detect the insurance fraud, data mining techniques have been applied most extensively. They have used logistic models, Neural networks, Bayesian belief Network and Decision Trees, which provide primary solution to the problems inherent in the detection and classification of fraudulent data. It also found gaps between FFD and the needs of industry to encourage additional research on neglected topics.
Alberto cano, Cristobal Romero, Amin Yousef, Mohammed Noaman, HabibmoussaFardoun, Sabastian Ventura	Classification Algorithms	Proposed prediction model to find out the performance of student and to decide which student to drop out in high schools.
R.Martin ,G. Thomas , K. Charles, O.Epitropaki, R.MC Namra	Data Mining Techniques	This paper identifies the relationship between subordinates and leader. The quality of relation can be predicted with the help of work related reactions. In order to find the better quality relation between managers and employees, data mining techniques can be used.
2011, Ernest H, OboyleJr, Ronald H, Humpherey Jeffery M, Pollack Thomas H, Hawer, Paul A.Story	Correlation Techniques	This article analysed relationship between Emotional intelligence and job performance. This Emotional intelligence whether affects the job performance or not. Classifying Emotional Intelligence in to 3 streams. Those streams could use correlation techniques to produce Five factor model. All the three streams have been demonstrated by using Dominance analysis.
2012, Brijeshkumar, Baradwaj, Saurabh pal	Classification Algorithm, ID3 Algorithm	To evaluate students performance, classification task has been used and also decision tree has been used as there are several approaches in data classification. By this task we can classify students either to dropout, need special attention, and to provide appropriate advising.
2010, James B- Avey, James L. Nimmicht, Nancy – Graber Pigeon	Data Mining Techniques	This paper examines the relationship between psychological factors and employee performance. This paper demonstrates that psychological capital is associated with multiple measures of employee performance.

CONCLUSION:

From these researches we can understand that Data mining is the predominant field in order to predict the employee performance and based on their performance we can suggest either the particular employee can continue in the same organization or he can quit and move to other organization. Finding weak employee and categorising types of employees with the help of data mining techniques has been done. Data mining techniques not only supports for predicting employee performance but also it helps in educational field to predict student performance. Few researches reveals that various skills to be developed for the fresh graduates to settle in appropriate organization. In some researches the relationship between the employee

performance and psychological factors and various attributes has been examined. In addition to that relationship between emotional intelligence and job performance has been determined by using data mining tools.

REFERENCE:

- [1] SujeetNarendra Mishra, DevRagavendraLama,"A Decision making model for Human Resource Management in organization using data mining and predictive analytics", International journal of computer science and information security(IJCSIS)Vol.14 No.5, May 2016
- [2] AkhileshK.Sharma, Dr.KamilijitLactaria," Data mining based prediction for employees skill enhancement using pro skill-Improvement program and performance using classifier scheme algorithm", International journal of advanced research in computer science, Vol.2,No.3, March 2013.

- [3] Hamidhajantan, mazidahputeh Abdul Razak Hamdan and Zuliahali Othman, "Applying Data Mining classification techniques for employee's performance prediction", 2010
- [4] jiechaochang, "Data mining research in education", March 2017.
- [5] Rahul edida, Rakshit Vahe, Rahul reddy, Rahul j, Abhilash, Deepti Kulkarni, "Employee attrition prediction", Management journal, 2018.
- [6] Kadir Eyasu Abdul Kadir, Fulea Amena Tolsa, "Predict and analysis of employee performance in bank using classification algorithms", International journal of interdisciplinary current advanced research (IJICAR), Vol. 1, No. 1, Feb 2019.
- [7] Ananya Sarkar, S.M. Shamim, Dr. Md. Shahiduz Zama, Md. Mustafizur Rahman, "Employees performance analysis and prediction using K means Clustering and decision tree algorithm", Global Journals, Vol. 18, No. 1, 2018.
- [8] Yuezaho, Maciej, Hryniewrcki, Francesca cheng, Boyang Fu, and Xiaoyuzhu, "Employee turnover prediction with machine learning: A reliable approach", January 2019.
- [9] Himanshu Kaumar Singh, K. Shitij Vishnavat, R. Srinivasan, "Employee performance and leave management using data mining techniques", International journal of pure and applied mathematics, Vol. 118, No. 20, 2018.
- [10] Subham Tupe, Chetan Mahajan, Dnyanshwar Plenchwar, Pratik Deo, "Employee performance evaluation system using ID3 Algorithm", International journal of innovative research in computer and communication engineering (IJRCCE), Vol. 5, No. 2, Feb 2017.
- [11] Pooja Thakar, Anil Mehta, Manisha, "Performance analysis and prediction in educational data mining: A Research Travalogue", International journal of computer applications, Vol. 110, No. 15, Jan 2015.
- [12] Farhad Shebani, Predicting the individual job satisfaction and determining the factors affecting it using CHAID Decision data mining algorithm", European Journal of Engineering research and science (EJERS) Vol. 4, No. 3, March 2019.
- [13] Nor Azziaty, Abdul Rahman, Kian Iamtan, and Chenkimlim, "Predictive Analysis and Data mining among the Employment of fresh graduates students in HEI", AIP Conference proceedings, 2017.
- [14] Subhamkarande, Ajay Shelake, Sivagami M, Sharon Sophia, "Prediction of Employee retention using Cassandra and Ensemble learning", International Journal of Recent technology and Engineering (IJRTE) Vol. 8, No. 1, May 2019.
- [15] Snikhan, Khera, Divya, "Predictive Modelling of Employee Turnover in Indian IT Industry using machine learning techniques", Journal published India, 2019.
- [16] Zarmina Jaffar, Dr. Waheed Noor, Zartash Kanwal, "Predictive Human Resource Analytics using Data mining classification techniques", International Journal of Computer, 2019.
- [17] Xiang Gao, Juhao Wen, Cheng Zhang, "An Improved Random Forest Algorithm for predicting Employee turnover", Hindawi, Mathematical problems of Engineering, 2019.
- [18] Amirah Mohamed Shahir, Wahidah Husain, Nuraini Abdul Rashid, "A review on predicting students performance using data mining", The third information systems international conference, 2015.
- [19] Saurabh Pal, "Evaluation of Teacher's performance: A Data mining Approach", January 2014.
- [20] Stefan Strohmeyer, Franca Piazza, "Domain driven data mining in human resource management: A review of current research", 2012.
- [21] EWT. Ngai, Yong Hu, Y.H. Wong, Yijunchen, Xinsun, "The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature", 2010.
- [22] Alberto cano, Cristobal Romero, Amin Yousef, Mohammed Noaman, Habibmoussa Fardoun, Sabastian Ventura, "Early dropout Prediction using data mining.
- [23] R. Martin, G. Thomas, K. Charles, O. Epitropaki, R. MC Namra, "The role of leader-member exchanges in mediating the relationship between locus of control and work reactions.
- [24] Ernest H. Oboyle Jr, Ronald H. Humphrey Jeffery M, Pollack Thomas H, Hawer, Paul A. Story, "The relationship between job performance and emotional intelligence.
- [25] Brijesh Kumar Bharadwaj, Saurabh Pal, "Mining Educational Data to analyse Students performance", January 2012.