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 is 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 large databases. Also, to identify previously hidden patterns in one step. Several types of data such as Relational database, data warehouse, Advanced DB and information repository, object oriented and object relational databases, transactional and spatial databases, heterogeneous and legacy database, multimedia and streaming data base, test 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 ≥ 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

<table>
<thead>
<tr>
<th>Category</th>
<th>Year, Author(s)</th>
<th>Methodology</th>
<th>Key Findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Survey of papers published in predicting</td>
<td>2010, Hamidahjantan, mazidahputeh, abdulrazah, ha, zulaihaalimdad otman</td>
<td>Classification Algorithms</td>
<td>Identified accuracy of classification technique in Data mining that helps to promote employee based on his performance.</td>
</tr>
<tr>
<td>predicting employee performance to promote him</td>
<td>2017, yecheochoeng</td>
<td>CRISP-DM cycle process</td>
<td>Identified Data Mining Technique to enhance decision making and analyzing new patterns and relationships for organizations.</td>
</tr>
<tr>
<td></td>
<td>2018, Rahul yedida, Rakhshahrhibulash, rahulreddy, rahul j, deetikulkarni</td>
<td>K-Nearest neighbors algorithm. It also includes Artificial neural network, Decision Tree, Logistic regression</td>
<td>Identified several methods of classification determined accurate results among different algorithms K – Nearest neighbor gives accurate results.</td>
</tr>
<tr>
<td></td>
<td>2019, KedirEyasn, AbdulKadhir, FlucaAmena, Toltsa</td>
<td>Data Mining Classification Algorithms, CRISP-DM techniques, Hybrid Data Mining process Model.</td>
<td>Efficiency and Effectiveness of employees is determined using Data Mining Classification Technique with basic parameters. It does not include confidential information about employee.</td>
</tr>
<tr>
<td></td>
<td>2018, Ananyasarker, SM, Shamim, Dr. M.D. Shahidul Zama and Md. Mostafizur Rahman</td>
<td>K- means Clustering, Decision Tree Algorithms.</td>
<td>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.</td>
</tr>
</tbody>
</table>

Survey of papers published in predicting Employee turnover

<table>
<thead>
<tr>
<th>Year, Author(s)</th>
<th>Methodology</th>
<th>Key Findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>2017, SubhamTupe, Chetan Mahajan, Dyanshwaruplenczwar, Pratek Deo.</td>
<td>Entropy, Decision Tree Algorithm, ID3 Algorithm.</td>
<td>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.</td>
</tr>
<tr>
<td>2019, Farhad Sheybani</td>
<td>CHAID Decision Tree algorithm</td>
<td>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.</td>
</tr>
<tr>
<td>2017, Nor Azziaty, Abdul Rahman, Kian lanitan, and Chenkimil</td>
<td>Supervised and unsupervised machine learning Algorithm such as K- Nearest Neighbor, Naive Bayes, Decision tree, Neural network, Logistic Regression, Support vector machine, CRISP-DM model</td>
<td>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.</td>
</tr>
<tr>
<td>2019, SubhamKarde, Ajay Shelake, sivagami M, Sharon Sophia.</td>
<td>Classifier techniques such as Support vector machine, Multi layer perceptron, Logistic Regression, Voting classifier, Apache Cassandra</td>
<td>Built an ensemble learning model which is a combination of Support Vector Machine, Logistic Regression, Random Forest. Based on accuracy, this model will able to predict the turnover of employees. Final classification is based on Weighted average.</td>
</tr>
<tr>
<td>2019, Snikhan, Khera, Divya</td>
<td>Supervised machine learning algorithm, Support Vector Machine.</td>
<td>The Employee data can be tested for this accuracy by using supervised machine learning classification models and hence validated.</td>
</tr>
<tr>
<td>2019, ZarminaJaffar, Dr. Waheed Noor, Zartash Kanwal</td>
<td>Data Mining Techniques such as J48, Naive Bayes and Logistic Regression.</td>
<td>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.</td>
</tr>
<tr>
<td>2019, Xiang Gao, Juhhao Wen, Cheng Zhang</td>
<td>Decision tree algorithm.</td>
<td>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.</td>
</tr>
<tr>
<td>2016, SujnetNarendramishraDevR</td>
<td>Data Mining Techniques</td>
<td>The necessity of Data mining Techniques in order to predict analytics for human resource management is discussed.</td>
</tr>
</tbody>
</table>
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:
Hamidhajantan, mazidahputehAbdulRazakHamdan and Zuliahali Othman, “Applying Data Mining classification techniques for employee’s performance prediction”, 2010


Brijesh Kumar Bharadwaj, SaurabPal,“Mining Educational Data to analyse Students performance”,January 2012.