International Journal of Advance Research in Science and Engineering
Vol. No.6, Issue No. 06, June 2017
www.ijarse.com
IJARSE
ISSN (O) 2319 - 8354
ISSN (P) 2319 - 8346

An efficient classifiers to predict performance of Employees

G.Devi[1], M.Prasanthi[2]

[1]*Assistant Professor, Dhruva Institute of Engineering & Technology, Hyderabad (India)*

[2]*Assistant Professor, Srisunflower college of Engineering & Technology, challapalli (India)*

ABSTRACT

The main objective of this paper is to identify the performance of Employees in an organisation by applying predictive data mining using classification based algorithms.There are various classification algorithms to check performance of Employees,such as *Multilayer Perception, Naïve Bayes, SMO, J48 and REPTree using WEKA an Open source tool. As a result, statistics are generated based on all classification algorithms and comparison of all five classifiers is also done in order to predict the accuracy and to find the best performing classification algorithm.*

Keywords: .Multilayer perception,Naive Bayes,EDM,SMO,J48,REPTree

I. INTRODUCTION

Data mining, also called Knowledge Discovery in Databases (KDD), is the field of discovering novel and potentially useful information from large amounts of data.here is the list of steps involved in the knowledge discovery process those are data cleaning,data integration,data selection,data transformation ,datamining,pattern evaluation knowledge presentation In any organization data mining field concentrate on prediction more often as compare to generate exact results for future purpose.Employee data mining (EDM) is the application of data mining technique on organization data.The objective of EDM is to analyze such organozation data,EDM deals with developing new methods to explore the employee data.Employee data mining uses many techniques such as Decision trees,Nueral networks,Naive Bayes, K-nearest neighbour and many others .

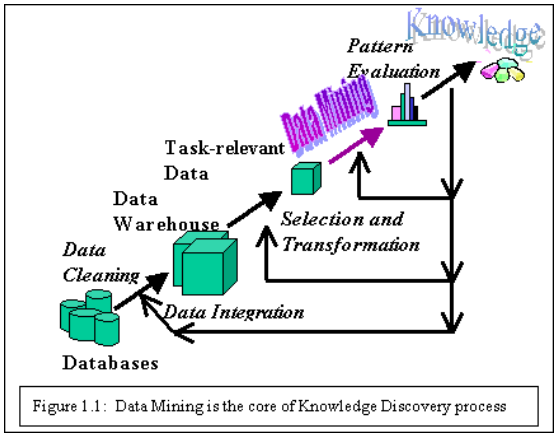

Figure 1.1: Data Mining is the core of Knowledge Discovery process

IJARSE
ISSN (O) 2319 - 8354
ISSN (P) 2319 - 8346

II. PROPOSED METHODOLOGY

A survey cum experimental Methodology is used a number of factors that are considered to have influence on the performance of a student are identified. These influencing factors are categorized as input variables. For this work, recent real world data is collected from high school. This data is then filtered out using manual techniques. Then data is transformed into a standard format required by the WEKA tool. After that, features and parameters selection is identified.

III. TOOLS AND TECHNIQUES

In this paper variety of data mining techniques are used for prediction of Employees in Employee Data mining the techniques are classification ,regression and density estimation during this work classification techniques for prediction are used the output dataset is tested and analyzed with five classification such as Multilayer Perception, Naïve Bayes, SMO, J48 *and* REPTree using WEKA an Open source tool.

IV. DECISION TREE INDUCTION

A Decision Tree is a flow chart like tree structures internal nodes denote a test on an attribute, Branches reprents out comes of tests leaf node hold class labels, Root node is the top most node. Decision tree classifiers are so popular because,the construction of a decision tree does not require any domain knowledge or parameter setting they can handle high dimensional data ,they have good accuracy. Decision may perform differently depending on the data set. these are used in various applications like medicine ,astronomy ,financial analysis.

V. SIMULATION CASE STUDY

A Total of 101 records are taken for the analysis of this selected high potential variables using select attributes facility of WEKA is done,chi squared attribute,info gain attribute and relief attribute evaluator are used to rank the variables ranker search method technique of WEKA is also applied.

The data set during this work is tested and analyze with five classification algorithms those are Multilayer Perception, Naïve Bayes, SMO, J48 and REPTree finally it has been investigated that J48 technique perform best with accuracy 75%.

Name of Classification	Class	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area
Algorithm							
Multilayer Perception	NQ	0.83	0.44	0.807	0.83	0.822	0.77
	Q	0.55	0.162	0.605	0.553	0.578	0.773
Naive Bayes	NQ	0.76	0.596	0.741	0.762	0.751	0.648
	Q	0.40	0.238	0.432	0.404	0.418	0.648
SMO	NQ	0.88	0.766	0.721	0.886	0.795	0.56
	Q	0.23	0.114	0.478	0.234	0.314	0.56
J48	NQ	0.81	0.574	0.761	0.819	0.789	0.713

IJARSE
ISSN (O) 2319 - 8354
ISSN (P) 2319 - 8346

	Q	0.42	0.181	0.513	0.426	0.465	0.713
REPTree	NQ	0.83	0.681	0.733	0.838	0.762	0.667
	Q	0.31	0.162	0.469	0.319	0.38	0.667

Mining technique	**Accuracy**
Multi layer perception	66%
Naive Bayes	65%
SMO	68%
J48	75%
REPTree	67%

VI. CONCLUSION AND FUTURE SCOPE

In this paper,classification techniques are used for prediction on the dataset of 101 employees to predict and analyze Employee performance among all data mining classifiers J48 performs best with 75%accuracy and therefore J48 proves to be potentially effective and efficient classifier algorithm.Hence the future of EDM is promising for further reaserch and can be applied in other areas like medicine,sports and share market due to the availability of huge databases.

REFERENCES

1. Zaïane, O. (2001)," Web usage mining for a better web-based learning environment", Proceedings Of Conference on Advanced Technology For Education, 60-64.
2. Zaïane, O. (2002), "Building a recommender agent for e-learning systems". Proceedings of the International Conference on Computers in Education,55–59.