ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Special Issue 1, March 2017

A Comparative Study on Analysis of Teachers' Performance using Classification Techniques

V. Shanmugarajeshwari[1], R. Lawrance[2]

Research Scholar, Department of Computer Science, Ayya Nadar Janaki Ammal College, Sivakasi, Tamil Nadu, India

Director, Department of Computer Applications, Ayya Nadar Janaki Ammal College, Sivakasi, Tamil Nadu, India

ABSTRACT:Data mining is used to extract the meaningful information from a large data source using some patterns and methods. It is one of the potential research field regarding interdisciplinary aspects. Educational data mining is one the emerging disciplines in the present scenario. One of the common tools to evaluate teachers' performance is the course evaluation questionnaire and based on the students' perception. Classification techniques in the data mining plays an important role in the field of educational data mining. The main goal of this research work is to predict the teachers' performance by using the relevant features. The proposed methodology consists of the phases like preprocessing, attribute selection, classification based on decision tree and performance evaluation. In the data preprocessing phase, the missing values have been removed. The attributes are transformed into a categorized format using the categorization process. Gain ratio, chi square and information gain feature selection techniques are applied on preprocessed data. The relevant attributes selected are predicted using classification techniques. Classification techniques are described and used for educational data mining. In this paper, it has been described that comparative study of classification techniques such as naïve bayes, decision tree, and C50. Among these the classification process is based on C5.0 algorithm with good classification accuracy.

KEYWORDS: Educational data mining; Performance appraisal; Teachers' performance; Feature selection; Decision tree algorithms; classification techniques.

I.INTRODUCTION

Data mining is used to extract the meaningful information from large data using some patterns. It has been used in many applications such as educational data mining, web mining, text mining and so on. It is an emerging discipline, concerned with developing methods for exploring the unique types of data that come from educational settings, and using those methods to better understand of academic system [1]. Methods are different from standard data mining methods. There are three main goals in educational data mining:
Pedagogical: to help in the design of enlightening contents and improvement on the academic performance of the students. Managerial: to optimize the organization and maintenance of education infrastructures, areas of interest and study researches.
Commercial: to help in students' recruitment in any private education [2]. This research focuses on pedagogical educational data mining.

II.LITERATURE REVIEW

Literature survey refers to a critical summary. Literature reviews contextualize research about a topic. A literature review is an evaluative report of studies found in the literature related to a selected area. The review should describe, summarize, evaluate and clarify this literature. It should give a theoretical basis for the research and help the researcher to the nature of the research [3]. It reviews that what have already been done in the framework of a topic. Therefore, on

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Special Issue 1, March 2017

the basis of the existing knowledge, everyone can build-up an innovative idea and concept for further research purpose [4].

Lawrance, R., *et al.* [5]present a paper on students' performance analysis using educational data mining is one of the emerging discipline which includes the process of analyzing the students' details using different attributes. Attributes such as students' name, roll number, previous semester marks, attendance, assignment, seminar performance, lab work and gender are used to evaluate the students' performance (Pass / Reappear). In this paper, classification techniques are described and used for educational data mining. The classification process is based on C5.0 algorithm with good classification accuracy. The system is helpful for the learners as well as to the teachers for the academic performance evaluation. It is a warning system for the students' to improve their study performance.

Agaoglu, M., [6] used four different classification techniques, decision tree algorithms, support vector machines, artificial neural networks, and discriminant analysis to build classifier models. Their performances are compared over a data set composed of responses to the students' questionnaire using accuracy, precision, recall, and specificity performance metrics. The main aim is to improve the instructor performance. In this paper C5.0 algorithm was used to predict and improve the instructor performance. This paper will help the firstly, effectiveness and expressiveness of data mining techniques, specifically decision tree algorithms, boosting, C4.5, SVM, ANN, and DA in higher educational mining are presented over a dataset from the daily life. Secondly, using the findings of the variable importance analysis for the classifiers, it is shown that there are many possible improvement areas in the design of the measurement instruments used in instructors' performance evaluation.

Shanmugarajeshwari., *et al.* [7] present data mining is widely used in educational field to extract useful information. It is the method of obtaining hidden, unknown and probably significant knowledge from huge voluminous amount of data. Educational data mining is one of the emerging discipline that can be applied for analyzing teachers' performance evaluation and students' performance evaluation. It has been found that educational data mining is the most frequently used assessment technique in higher education to know how the courses are taught. In this survey, educational data mining is applied to evaluate teachers' performance and students' performance and find out the interesting information to credit teachers' appraisal.

Ughade, P., *et al.* [8] present a paper on faculty performance that has calculated on the basis of two parameters (i.e.) Student's feedback and the result of student in that subject. In existing system they define two approaches one is multiple classifier approach and the other is a single classifier approach and comparing them, for relative evaluation of faculty performance using data mining techniques. At multiple classifier approach K-nearest neighbor (KNN) is used in first step and Rule based classification is used in the second step of classifying while in single classifier approach only KNN is used in both steps of classification.

From the literature review, it is inferred that teachers' performance can be evaluated by applying, data mining techniques, classification, fuzzy logic and overall performance can be evaluated by classification techniques. Based on the literature survey our research work classified as three phases. They apply for this research work.

Phase I: Preprocessing
- i) Dataset Description
- i) Removing Missing Values
- ii) Categorization

Phase II: Feature Selection
- i) Chi-Square Feature Selection
- ii) Information Gain Feature Selection
- iii) Gain Ratio Feature Selection

Phase III: Classification
- i) Naïve Bayes
- ii) Decision Tree

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Special Issue 1, March 2017

iii) C50

III.DATA SET DESCRIPTION

A. *Data Collection*

The input data have been collected from Ayya Nadar Janaki Ammal College, Sivakasi from the students and the staffs of computer applications department (MPhil, MCA, BCA, Staffs and Students). The data size is 598 records X 48 attributes.

B. *Data Preparation*

The students' data are collected from evaluation of the teacher by students for the academic year 2016-17. The staffs' data are collected from teachers' questionnaire forms. Students' frank and unbiased evaluation of the teacher/teaching helps the teacher as well as the department to improve teaching – learning process. Given below are some points / aspects of evaluation. Selected the appropriate grade in the box (E - Excellent, VG – Very Good, G – Good, F – Fair, P – Poor). Attributes used in this study are described in Table I and Table II.

TABLE I. DATA SET INFORMATION

Attributes	Description	Possible Values
PTC	Punctuality to the class	E, VG, G, F, P
LP	Lesson planning	E, VG, G, F, P
EO	Expression of objectives	E, VG, G, F, P
SOSM	Systematic organization of subject matter	E, VG, G, F, P
PSC	Preparation of the subject for the classes	E, VG, G, F, P
PHSSM	Provision of handouts / Synopsis / Study materials	E, VG, G, F, P
ESS	Encouragement of self-study	E, VG, G, F, P
CL	Clarity of the lecture	E, VG, G, F, P
LWD	Legibility of writing & drawing	E, VG, G, F, P
EOD	Explanation of the concepts	E, VG, G, F, P
UMICTEL	Use of modern techniques / ICT / E lessons	E, VG, G, F, P
MRS	Meeting the requirements of students	E, VG, G, F, P
ACS	Assistance and counselling to students	E, VG, G, F, P
MS	Motivation of students	E, VG, G, F, P
ISET	Interaction with students for effective thinking	E, VG, G, F, P
PSAQ	Permitting students to ask questions	E, VG, G, F, P
CDRS	Clearing the doubts raised by students	E, VG, G, F, P
ESO	Encouraging the students to be original	E, VG, G, F, P
ITS	Impartiality towards students	E, VG, G, F, P
MDC	Maintenance of discipline in the class	E, VG, G, F, P

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Special Issue 1, March 2017

COS	Coverage of syllabus	E, VG, G, F, P
ASS	Acceptance of suggestions from students	E, VG, G, F, P
CGD	Conduct of group discussions	E, VG, G, F, P
LWIPSA	Lab work instruction / Problem solving ability	E, VG, G, F, P
EILELPPP	Encouragement of innovative learning, E-learning, power-point presentation	E, VG, G, F, P
TEPEPIAL	Teacher explains the practical / experiments and pays individual attention in lab.	E, VG, G, F, P
PCQ	Periodic conduct of quizzes	E, VG, G, F, P
VAS	Valuation of answer scripts	E, VG, G, F, P
CLWC	Correction of lab work / composition	E, VG, G, F, P
CSA	Conduct of seminars / assignment	E, VG, G, F, P
EASST	Evaluation of answer scripts within stipulated time	E, VG, G, F, P
SIBE	Suggestions given for improvement based on evaluation	E, VG, G, F, P
STPT	Suggesting the topic / problem in time	E, VG, G, F, P
PRM	Provision of reference materials	E, VG, G, F, P
PRW	Periodical review of the work	E, VG, G, F, P
CPRT	Consolidation of project report in time	E, VG, G, F, P
EPPS	Encouragement for paper presentations in seminars	E, VG, G, F, P
PMP	Preparation of manuscripts for publications	E, VG, G, F, P

It will be categorized into 5 groups, such as Excellent – 0.95, Very Good – 0.85, Good – 0.75, Fair – 0.65, Poor – 0.30.

Here, all variables are in the form of categorical values. The result is the dependent variable and all other variables are the predictor variables.

TABLE II. TEACHERS' QUESTIONNAIRE DATA SET

Attributes	Description	Possible Values
Q1	How will you compensate students' learning with your absenteeism?	100%, >75%, >50%, Not following Blooms' Taxonomy
Q2	How will you compensate students' learning with your absenteeism?	Substitution class to take concerned topic, Extra hour teaching,

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Special Issue 1, March 2017

		Combined class, Adjustment class
Q3	How will you effectively manage a class with troubled students?	Give some extra work to them, Listening students in one to one meeting, Interactive session, Well-designed enjoying lesson
Q4	Communication skills	Very High, High, Average, Low
Q5	How effectively you are using ICT in your classroom?	Very High, High, Average, Low
Q6	How will you handle a weak student?	Motivate the student, Confront individually with the student about the problems he / she is facing, Adopt different teaching methods, Realize strength and its capability
Q7	How much support does the administration at this college give to the teaching staff?	A great deal, A lot, A moderate amount, A little , None at all
Q8	How well do teachers at this college collaborate with each other?	Extremely well, Very well, Somewhat well, Not so well, Not at all well
Q9	How much attention does this college give to your professional growth?	A great deal, A lot, A moderate amount, A little , None at all
Q10	Overall, are you satisfied with the teaching experience at this college, neither satisfied nor dissatisfied with it, or dissatisfied with it?	Extremely satisfied, Moderately satisfied, Slightly satisfied, Neither satisfied nor dissatisfied, Slightly dissatisfied, Moderately dissatisfied, Extremely dissatisfied

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Special Issue 1, March 2017

C. PREPROCESSING

The students' data have been taken from preprocessing step. During preprocessing all individual tables are combined into a single table formed with all sufficient data and also missing values are also omitted. Preprocessing can increase the classification accuracy. The proposed system of preprocessing contains two steps.

Step 1: Removing missing values

Step2: Categorization

IV.METHODOLOGY

Now-a-days, education data mining plays a major role in the society. In this paper, educational data mining was used to improve the teachers' performance using feature selection and classification techniques. Teachers' performance evaluation for a college has been done, in order to improve teachers' performance by decision tree in classification technique. Educational data to predict the teachers' performance as well as students' performance in return improve academic performance.

A. FEATURE SELECTION

For feature selection, a number of techniques are available in data mining some of them are chi-squared feature selection (CFS), information gain feature selection (IGFS), gain ratio feature selection (GRFS) and correlation based feature selection (CBFS). Feature selection is also known as attribute selection [9]. In selecting the relevant attributes and discarding the irrelevant attributes, various feature selection techniques were applied to the preprocessed data set which has 598 samples, concentrated on pick out the top attributes

This research focuses on three feature selection methods are compared. The best feature selection method is gain ratio feature selection have been applied to the preprocessed data. It is a modification of the information gain that reduces its bias. It takes the number and size of branches into account when choosing an attribute.

B. GAIN RATIO FEATURE SELECTION

One of the best feature selection method is gain ratio feature selection. The purpose of feature selection method is to extract the relevant features and discard the irrelevant features [9]. In gain ratio feature selection, the subset has been selected using entropy D_j value (1) and information gain $_A(D)$, (3). Gain ratio (4) is calculated using the equations (1) and(3)

$$\text{Entropy } (D_j) = -\sum_{j=1}^{m} pj \log 2(pj) \quad (1)$$

$$\text{Info Gain } (D, A) = \text{Entropy } (D_j) - \sum_{j=1}^{v} \frac{Dj}{D} * \text{Entropy } (Dj) \quad (2)$$

$$\text{Gain Ratio } (A) = \text{Entropy } (D) - \text{Information Gain }_A (D) \quad (3)$$

$$\text{Information Gain }_A (D) = \sum_{j=1}^{v} \frac{Dj}{D} * \text{Entropy } (Dj) \quad (4)$$

From the above equations, the gain ratio subset selector has been produced. It is used to improve the classification accuracy because it extracts the relevant attributes only. Here, redundant attributes are removed.

TABLE III. COMPARATIVE ANALYSIS ON FEATURE SELECTION METHODS

Feature Selection Methods	Selected Attributes()
Chi-squared feature selection	6
Information gain feature selection	6
Gain ratio feature selection	5

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Special Issue 1, March 2017

Fig. 1 methodology framework uses two phases: The removing missing values are removed in the preprocessing phase. The attributes are transformed into a categorized format using categorization process phase.

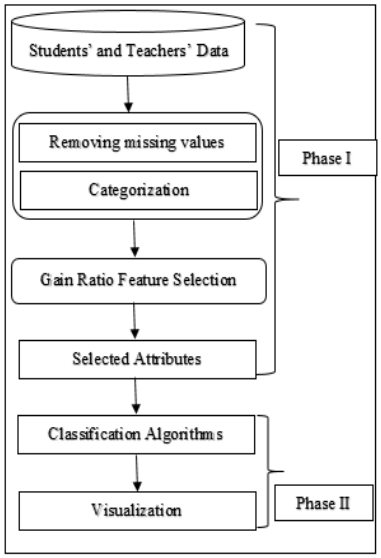


Fig. 1. Methodology Framework

V. CLASSIFICATION

Classification is the process to classify the data objects with the predetermined class labels. Classification is one of the supervised machine learning algorithms [10]. In classification, the preprocessed dataset has been divided into two sets, such as training set and a test set. Using the training sets the classifier model has been developed. The test set was applied to the classification model, and it will be classified.

TABLE IV. COMPARATIVE ANALYSIS ON CLASSIFICATION

Algorithms	Accuracy (%)
Decision Tree	91.6
CHAID	60.5
C50	92.3
ID3	87
C4.5	67.77
CART	89.9
Naïve Bayes	86
K-NN	78
SVM	91.3
ANN	91
DA	90.5

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Special Issue 1, March 2017

TABLE V. PERFORMANCE ANALYSIS OF CLASSIFICATION

Classification Methods	Accuracy (In %)	Time (In Sec)
Naïve Bayes	86	0.24
Decision Tree	91.6	0.50
C50	100	0.18

A. *DECISION TREE*

A decision tree is a simple structure where each non-terminal node represents a test or decision on the considered data item. A decision tree can be used to classify an instance by starting at the root of the tree and moving through it until a leaf node [11].

Decision tree is a supervised classification, which predicts both the classifier and regression models. Classification trees are mainly used to classify an object to a predetermined class based on the attributes. The tree contains no incoming edge is called as root, A node with one outgoing edge is called as internal node, all other nodes are known as Leaf node which has no outgoing edge. Using the training sets the classifier model have been developed, testing set was applied to the classification model to predict the previously unknown class. There are many specific decision tree algorithms. They described below:

B. *CHI-SQUARED AUTOMATIC INTERACTION DETECTION(CHAID) ALGORITHM*

CHAID is one of the types in decision tree classification technique. This algorithm used for Prediction as well as for the detection of interaction between variables. The output generated by CHAID classification algorithm is highly visual and easy to interpret.

C. *C5.0*

C5.0 is one of the important decision tree classification algorithms. It can handle continues and categorical values. It can handle numeric attributes. Attribute selection is the fundamental step to construct a decision tree. Entropy, information gain and gain ratio are used to process attribute selection, using attribute selection C5.0 algorithm select which attribute will be selected to become a node of the decision tree and so on. Comparing with ID3 and C4.5, C5.0 is the highest speed. Pre-pruning was performed. One of the decision tree classification algorithms. Entropy, information gain and gain ratio measures are calculated and classification model was developed. Applied on the data set to determine the unknown samples.

Advantages:

i) It provides the accurate result
ii) It needs less memory space for large data set
iii) It takes less time to build a model
iv) It supports Boosting
v) Compare with ID3 and C4.5 it has the highest speed
vi) It can handle continuous value, date and categorical values. And easily handled the multi-value attributes and missing attribute from the data set

Disadvantages:

i) Empty branches and insignificant branches are allowed

D. *ID3*

ID3 stands for iterative dichotomiser3 developed by Quinlan (1986). ID3 algorithm works with the concepts of Information gain by calculating entropy of an attribute. It cannot deal with missing values and it has lowest speed. No pruning was performed. ID3 algorithm works with the concepts of information theory by calculating the entropy of an attribute. ID3 decision tree classification is used for selecting the appropriate attributes.

Advantages

i) It provides more accurate result compared than C4.5
Algorithm.
ii) ID3 algorithm generally uses categorical attributes for classification.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Special Issue 1, March 2017

iii) Missing values of an attribute is not allowed.
iv) It consumes less memory space.

Disadvantages

i) It requires long searching time.
ii) It takes more memory space than the C4.5 to large data set.

E. C4.5

C4.5 is one of the decision tree classification algorithms. It was developed by Quinlan (1993). It uses a gain ratio as a splitting criteria. By calculating entropy and splitting information of an attribute. It can handle numeric attributes and missing values. Faster than ID3 algorithm. It also cannot deal with missing values.

Decision tree is constructed by examining a set of training samples whose class labels are known. These features of known samples are applied in order to determine the properties of unknown samples

Advantages:

i) It provides the accurate result
ii) It needs less memory space for large data set
iii) It takes less time to build a model
iv) It has short searching time

Disadvantages

i) Empty branches and insignificant branches are allowed
ii) Over fitting is one of the most important problems in C4.5 algorithm

F. CART

Classification and regression trees developed by Breiman (1984). It is characterized by the fact it constructs binary trees, namely each internal node has exactly two outgoing edges. The splits are selected using towing criteria. Cost, complexity pruning is the stopping criteria. It has the average speed.

The Naïve bayes technique does not handle continuous data, dividing the continuous values into ranges could be used to solve this problem, but the division of the domain into ranges is not an easy task. In K-NN (K-Nearest Neighbors) algorithm, each tuple to be classified must be compared to each element in the training data, if there are q elements in the training set, this is O (q). Given n elements to be classified, this becomes an O (nq) problem. Given that the training data are of a constant size (although perhaps quite large), this can be viewed as an O (n) problem. Comparing with naïve bayes, K- NN (K-Nearest Neighbors), decision tree algorithms. Decision tree algorithms are suitable for EDM. Hence, it has been chosen C5.0 algorithm for classification.

VI. EXPERIMENTAL RESULTS

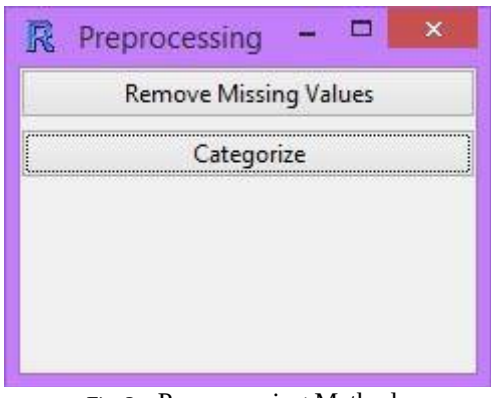


Fig. 2. Preprcoessing Methods

Fig. 2 shows the preprocessing methods.

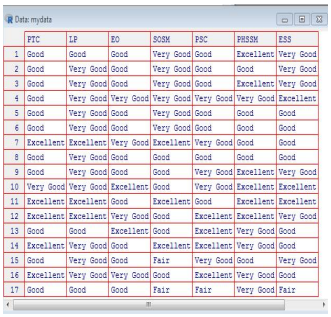
ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

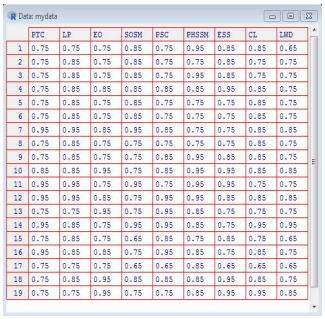
Website: www.ijircce.com

Vol. 5, Special Issue 1, March 2017



	PTC	LP	EO	SOSM	PSC	PHSSM	ESS
1	Good	Good	Good	Very Good	Good	Excellent	Very Good
2	Good	Very Good	Good	Very Good	Good	Good	Very Good
3	Good	Very Good	Good	Very Good	Good	Excellent	Very Good
4	Good	Very Good	Very Good	Very Good	Very Good	Very Good	Excellent
5	Good	Very Good	Good	Very Good	Good	Good	Good
6	Good	Very Good	Good	Very Good	Good	Good	Good
7	Excellent	Excellent	Very Good	Excellent	Very Good	Good	Good
8	Good	Very Good	Good	Good	Good	Good	Good
9	Good	Very Good	Good	Good	Very Good	Excellent	Very Good
10	Very Good	Very Good	Excellent	Good	Very Good	Excellent	Excellent
11	Excellent	Excellent	Good	Excellent	Good	Excellent	Excellent
12	Excellent	Excellent	Very Good	Good	Excellent	Excellent	Very Good
13	Good	Good	Excellent	Good	Excellent	Very Good	Good
14	Excellent	Very Good	Good	Excellent	Excellent	Very Good	Good
15	Good	Very Good	Good	Fair	Very Good	Good	Very Good
16	Excellent	Very Good	Very Good	Good	Excellent	Very Good	Good
17	Good	Good	Good	Fair	Fair	Very Good	Fair

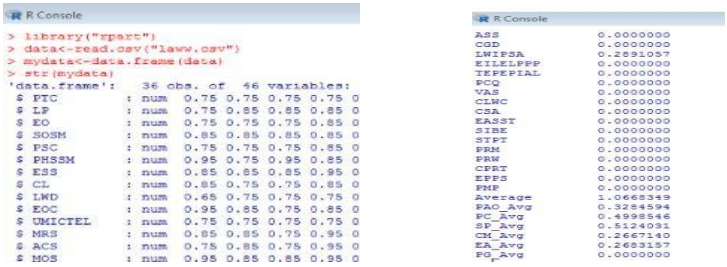
Fig. 3. Raw Data Set,



	PTC	LP	EO	SOSM	PSC	PHSSM	ESS	CL	LMD
1	0.75	0.75	0.75	0.85	0.75	0.95	0.85	0.85	0.65
2	0.75	0.85	0.75	0.85	0.75	0.75	0.85	0.75	0.75
3	0.75	0.85	0.75	0.85	0.75	0.95	0.85	0.75	0.75
4	0.75	0.85	0.85	0.85	0.85	0.85	0.95	0.85	0.75
5	0.75	0.85	0.75	0.85	0.75	0.75	0.75	0.85	0.75
6	0.75	0.85	0.75	0.85	0.75	0.75	0.75	0.85	0.75
7	0.95	0.95	0.85	0.95	0.85	0.75	0.75	0.85	0.85
8	0.75	0.85	0.75	0.75	0.75	0.75	0.75	0.85	0.75
9	0.75	0.85	0.75	0.75	0.85	0.95	0.85	0.85	0.75
10	0.85	0.85	0.95	0.75	0.85	0.95	0.95	0.85	0.85
11	0.95	0.95	0.75	0.95	0.75	0.95	0.95	0.75	0.75
12	0.95	0.95	0.85	0.75	0.95	0.95	0.85	0.85	0.85
13	0.75	0.75	0.95	0.75	0.95	0.85	0.75	0.75	0.75
14	0.95	0.85	0.75	0.95	0.95	0.85	0.75	0.95	0.95
15	0.75	0.85	0.75	0.65	0.85	0.75	0.85	0.75	0.65
16	0.95	0.85	0.85	0.75	0.95	0.85	0.75	0.85	0.75
17	0.75	0.75	0.75	0.65	0.65	0.85	0.65	0.65	0.65
18	0.75	0.85	0.95	0.85	0.85	0.85	0.95	0.85	0.75
19	0.75	0.75	0.95	0.75	0.75	0.85	0.95	0.95	0.85

Fig.4.Categorized Data

Fig. 3 shows the raw data. Raw data contains errors, missing values, inconsistent data and also it contains numerical data and categorical data. Fig. 4 raw data contains categorical data and also numerical data. It needs to be in generalized format. For that numerical data should be transformed into categorical data. In categorization all the attributes are transformed into a categorized format using certain conditions. Fig. 4 displays the categorization process.



```
R Console
> library("rpart")
> data<-read.csv("laww.csv")
> mydata<-data.frame(data)
> str(mydata)
'data.frame':   36 obs. of  46 variables:
 $ PTC       : num  0.75 0.75 0.75 0.75 0
 $ LP        : num  0.75 0.85 0.85 0.85 0
 $ EO        : num  0.75 0.75 0.75 0.85 0
 $ SOSM      : num  0.85 0.85 0.85 0.85 0
 $ PSC       : num  0.75 0.75 0.75 0.85 0
 $ PHSSM     : num  0.95 0.75 0.95 0.85 0
 $ ESS       : num  0.85 0.85 0.85 0.95 0
 $ CL        : num  0.85 0.75 0.75 0.85 0
 $ LWD       : num  0.65 0.75 0.75 0.75 0
 $ EOC       : num  0.95 0.85 0.75 0.85 0
 $ UMICTEL   : num  0.75 0.75 0.75 0.75 0
 $ MRS       : num  0.85 0.85 0.75 0.95 0
 $ ACS       : num  0.75 0.85 0.75 0.95 0
 $ MOS       : num  0.95 0.85 0.85 0.95 0
```

```
R Console
ASS             0.0000000
COD             0.0000000
LWIPSA          0.2891057
EILELPPP        0.0000000
TEPEPIAL        0.0000000
PCQ             0.0000000
VAS             0.0000000
CLWC            0.0000000
CSA             0.0000000
EASST           0.0000000
SIBE            0.0000000
STPT            0.0000000
PRM             0.0000000
PRW             0.0000000
CPRT            0.0000000
EPPS            0.0000000
PMP             0.0000000
Average         1.0668349
PAO_Avg         0.3284594
PC_Avg          0.4998546
SP_Avg          0.5124031
CM_Avg          0.2667140
EA_Avg          0.2683157
PG_Avg          0.0000000
```

Fig.5.Objects and Atttributes Data

Fig. 5 represents the data frame 36 objects of 46 variables. Table VI represents how the gain ratio feature selection have been applied to the preprocessed data. The relevant attributes only selected for further process.

TABLE VI. SELECTED ATTRIBUTES

Method	Selected Attributes				
Gain Ratio	PSC	Average	PAO_Avg	PC_Avg	SP_Avg

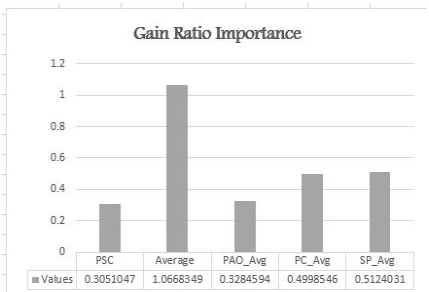


Fig.6.Gain Ratio Importance

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Special Issue 1, March 2017

Fig. 6 represents the gain ratio importance values to select the specific attributes. The selected attributes using the chi-square feature selection method which is shown in Table VII.

TABLE VII. SELECTED ATTRIBUTES

Method	Selected Attributes					
Chi_Square	PSC	LWIPSA	Average	PAO_Avg	PC_Avg	SP_Avg

Fig. 7 describes chi square feature selection based on observed frequency and expected frequency chi square filter has been calculated then, used to select the relevant attributes.

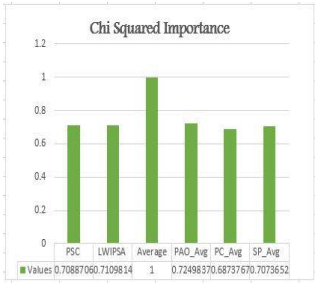


Fig.7.Chi-Square Importance,

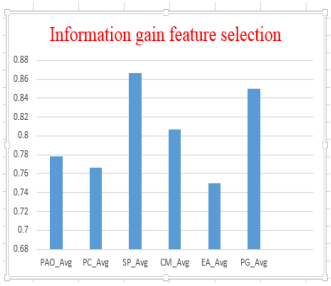


Fig.8.Information gain feature selection

Fig. 8 illustrates information have been gained from the entropy. Information gain feature selection method is based on the information gain and gain ratio these are the selected attributes using the information gain feature selection method.

The selected attributes using the information gain feature selection method which is shown in Table VIII.

TABLE VIII. SELECTED ATTRIBUTES

Method	Selected Attributes					
Information Gain	PAO_Avg	PC_Avg	SP_Avg	CM_Avg	EA_Avg	PG_Avg

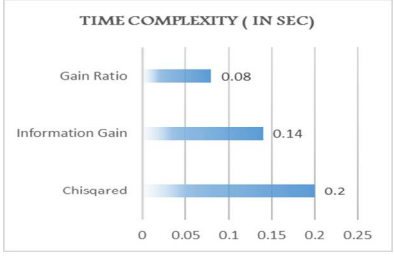


Fig.9.Time Complexity of three Feature Selection Methods Time Complexity (in milli sec)

Fig. 9 shows the feature selection using chi-squared method takes very long time (0.20 Milliseconds), the information gain method takes (0.14 Milliseconds), and gain ratio method takes very less time (0.08 Milliseconds).

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Special Issue 1, March 2017

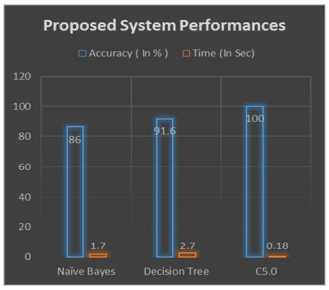


Fig.10.Proposed System Performance

Fig. 10 illustrates performance system performance using Naïve bayes, Decision tree and C50 accuracy.

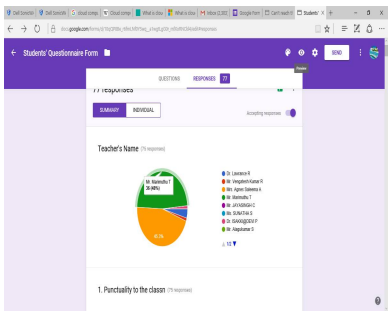
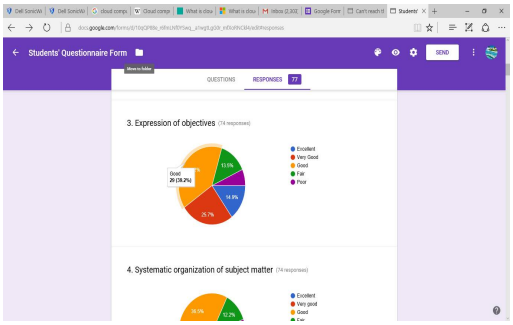


Fig.11.Google forms analysis dataset results

Fig. 11 shows the teachers' and students' questionnaire dataset collected by using google forms.



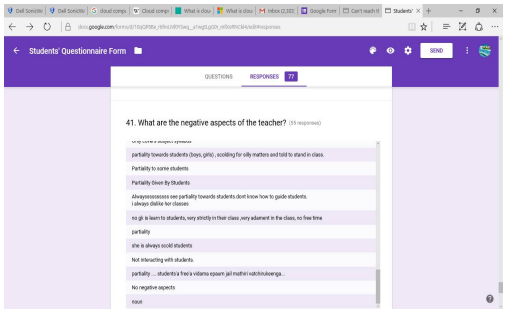


Fig.12.Google forms using analysis teachers' performance positive and negative results

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Special Issue 1, March 2017

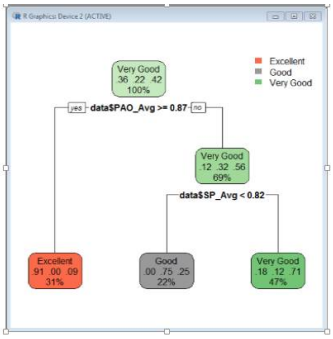


Fig.13.Teacher Performance,

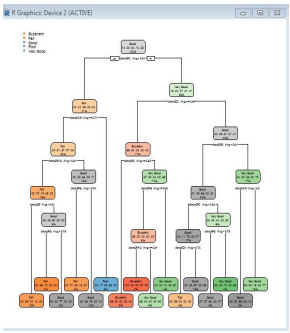


Fig.14.MCA Department Overall Teachers' Performance

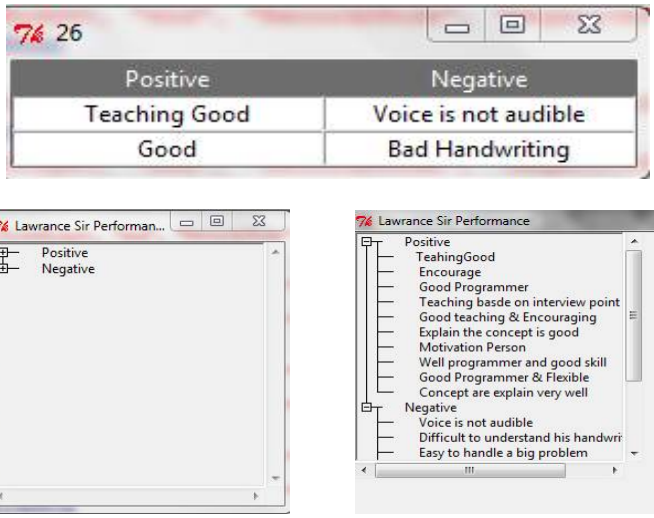


Fig.15.Teacher Performance, Positive and Negative

VII. CONCLUSION

Classification techniques are used to make better decisions in the educational data mining. In this, teachers' performance of educational data mining has been done using different approaches. Prediction is based on the teacher's performance evaluation and student performance evaluation. The primary aim of this proposed research is to motivate the teachers to improve their work performance, which can be used for judging purposes in order to make good administrative decisions. Data mining techniques show automated discovery of new associations and dependencies of attributes in the observed data. The C5.0 classification algorithm has 100% classification accuracy. The proposed algorithm is compared with benchmark algorithms such as decision tree induction and naïve bayes. In this research technique of data mining in education section is used to improve the teacher's performance.

REFERENCES

[1] Bakar, R.S., and Yacef, K, "The state of educational data mining in 2009: A review and future visions." JEDM-Journal of Educational Data Mining 1.1 (2009):3-17
[2] Barracosa, J.I.M.S.2011. Mining Behaviors from Educational Data
[3] http://library.queensu.ca/webedu/grad/Purpose_of_the_Literature_ Review.pdf
[4] http://ar.cetl.hku.hk/am_literature_reviews.htm

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Special Issue 1, March 2017

[5] Lawrance, R., Shanmugarajeshwari, V., "Analysis of Students' Performance Evaluation using Classification Techniques." IEEE International Conference on Computing Technologies and Intelligent Data Engineering(ICCTIDE'16), 10.1109/ICCTIDE.2016.7725375, 16426971
[6] Agaoglu, M., IEEE Translation and Content Mining, "Predicting Instructor Performance Using Data Mining Techniques in Higher Education." Department of Computer Engineering, Marmara University, Istanbul 334722, Turkey. Volume.4, 2016, pp:2379-2387
[7] Shanmugarajeshwari, V., Lawrance, R., "Teachers' Performance on Educational Data Mining: A survey." JOURNAL OF MANAGEMENT SCIENCE AND HUMANITIES (JOMSAH) BI-ANNUAL, ISSN: 2395-0625 Vol. 3 No. 1 June 2016, pp: 51-56
[8] Ughade, P., Mohod, W.S., "A Survey on Analysis of Faculty Performance Using Data and Opinion Mining." International Journal of Innovation Research in Computer and Communication Engineering (IJIRCCE), Volume. 4, Issue. I, January 2016, pp: 87-91
[9] http://en.wikipedia.org/wiki/Feature_selection
[10] http://www.saedsayad.com/classification.htm
[11] Quinlan, J. R. "Induction of decision trees." Machine learning 1.1 1986, pp. 81-106