Available Online at www.ijcsmc.com

International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

ISSN 2320–088X
IMPACT FACTOR: 6.017

IJCSMC, Vol. 6, Issue. 5, May 2017, pg.405 – 417

PREDICTION OF MU STUDENT'S PERFORMANCE USING DATA MINING TECHNIQUE

HEMANT SHARMA[1], SHIV KUMAR[2]

[1]M.Tech Scholar, Computer Science & Engineering, Mewar University, Gangrar, Chittorgargh-312901, India
hemantodichaya7@gmail.com
[2]Computer Science & Engineering, Mewar University, Gangrar, Chittorgargh-312901, India
Shivkumar004@gmail.com

Abstract*— The role of technology in education industry is increasing day by day specially after reaching the growth rate of the IT industry to the saturation point. That is why, even AICTE has closed the affiliation of engineering colleges in number of states. So, education industry can change their plan to survive in this environment by using data mining techniques to identify the outstanding students as well as to pay the extra attention to the poor performer students to improve their performance. As a Mewarian, this is our first duty to develop such type of system which can predict the performance of the MU students by learning their past results in terms of 10th marks, 12th marks, previous semester marks, current semester PCA marks and attendance record using data mining classification techniques of Naïve Bayes classification and J48 classification techniques.

***Keywords*— Mu, Data mining, Prediction, Naive ayes, J48, Performance, Classification**

I. INTRODUCTION

The biggest challenges in front of the education industry is to provide the job to each student of the colleges or university in any how situation to survive when students are not skilled and industry requires multi talented students those can work without training. The branded engineering colleges can survive in this scenario by selecting creamy layer students who can pay fee for the any type of training by using third party. But, Mewar University like university or colleges are unable to provide such type of services because they depend on the government scholarship policies to serve the nation by providing the platform to the poorest to the poorest

students of this country, the goal and objective is excellence one while the services are also very high class but not as par the requirement of the students. That is why performance or growth of the university decreasing day by day. As par last two years trends in admission of Faculty of Engineering & Technology School, this department will be closed in 2k18 until unless university will take precaution measures to survive because this school is the heart of the university or Chittorgarh in the field of education.

This type of problem can be solved by using latest technology like data mining. Data mining is like mining industry, only difference is that mining industry extracting ores from raw material while data mining techniques extracting useful data or information from the raw data. Data mining solves problem by analysing large amount of available data by providing useful pattern and rules using some classification method in following steps as shown in figure1:

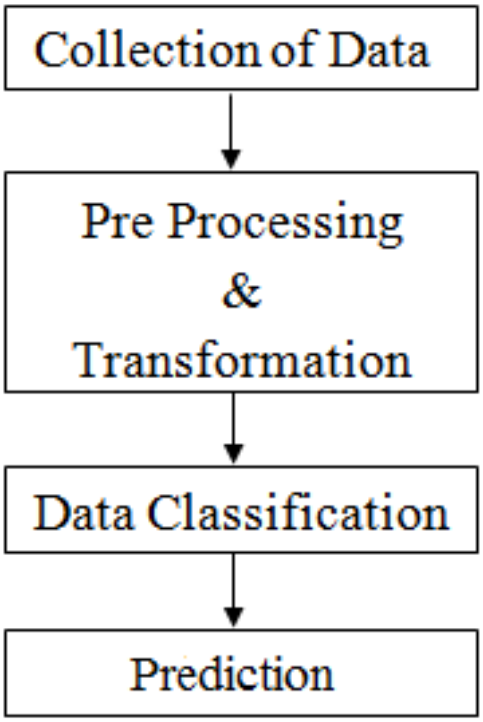

Figure1: Data mining steps

As university has large amount of data of their students since its existence. But the problem is not the storing data but the extracting the meaningful data from the available data. The problem of bad performance in the examination can be solved by deploying data mining tools to trace the record of the students and future assistance for their improvement. Our main aim is to identify different factors that affect the performance of the students by implementing data mining techniques. There are several data mining techniques. Some of them are:

- Primary techniques
- Secondary techniques

A primary technique includes the following techniques:

- Classification techniques
- Clustering techniques
- Association Rules techniques

A secondary technique includes the following techniques:

- Sequential Patterns techniques
- Regression techniques
- Deviation Detection techniques

We have used classification techniques in the proposed system that is why we are discussing only this technique. In classification techniques two sets of data are used:

- Training set data
- Test set data

Training set is a collection of given records, where each record has number of attributes. Attributes is the name of column id data is in tabular format. While each attribute is known as the class. It is used to build the model. Test set is also a collection of given records but it is used to validate the model. Generally number of records in training set and test are in the ratio. There are several classification techniques, some of them are:

- Decision Tree based Methods
- Rule-based Methods
- Memory based reasoning
- Neural Networks
- Genetic Algorithms
- Naïve Bayes and Bayesian Belief Networks
- Support Vector Machines

II. LITERATURE REVIEW

R.sumitha,et.al., [2016] [1], "Prediction of students outcome using Data mining Techniques" has discussed about the classification panel enables the user to apply classification and regression algorithms to the resulting dataset, to estimate the accuracy of the resulting predictive model, and also to visualize SMO, J48, REP TREE.

Abeer Badr El Din Ahmed,et.al., [2014] [2], "Data mining: A prediction for students‟ performance using classification method "has discussed the decision tree method on student's database to predict the student's performance on the basis of student's database. This study helps the student's to improve their performance and to identify student's those needed special attention to reduce failing rate and taking appropriate action at right time.

Dorina kabakchieva,et.al., [2013] [3] ,"Predicting students‟ performance by using data mining methods for classification has discussed to find out the class variable using the explanatory variable .it is possible to predicate.

Marie Bienkowski,et.al., [2012] [4], "Enhancing teaching and learning through Educational data mining and learning analytics" has discussed that higher Education institutions are applying learning analytics to improve the services. this is useful in majoring and improving grades.

Puja Thakur,et.al., [2015] [5], "Performance analysis and prediction in educational Data mining : A research Travelogue " has discussed about the challenging higher education facing today in making students knowledgeable and skill .

Dr.Mohd Maqsood Ali,et.al., [2013] [6], "Role of data mining in education sector" has describe the profile of successful and unsuccessful students based on GPA achieved during the semesters. It can also be used for dropout students, academic performance, teachers‟ performance, and students complaints by using IF THEN rule.

M.Ramaswami, et. al. [2010][7], " A chaid based performance prediction model in educational data mining has discussed the prediction model of students by using seven class predictor variable.

Arockian,et.al., [2011] [8], "Deriving association between urban and rural students programming skills" has discussed about FP tree and K-means clustering technique for finding similarity between urban and rural students programming skills. FP tree mining is applied to sieve the patterns from the dataset. K-means clustering is used to determine the programming skills of the students.

Komal S.Sahedani,et.al., [2013] [9], "A review: Mining educational data to forecast failure of engineering students has discussed that community colleges and universities can build model that predict high degree of accuracy by using clustering technique. By acting on these predictive models, educational institutions can effectively address issues reses from transfers and retention..

III. PROBLEM STATEMENT

Predicting students" performance becomes more challenging due to the large volume of data in educational databases. This is due to main two reason. First, the study of existing prediction methods is still insufficient to identify the most suitable methods for predicting the performance of students in the University. Second is due to the lack of investigations on the factors affecting student's achievements in particular courses. Therefore, a necessary literature reviews on predicting student performance by using data technique is

IV. OBJECTIVE

The main object of this research is as following:

- Study of Data mining techniques
- Prediction of Mewar University result using Naïve Bayes & J48 algorithm
- Implementation of these algorithms on Weka tool following parameter:
  - 10^{th} (TP) and +2(TP+2)
  - Attendance(AP)
  - Teacher assessment(TA)
  - Pre Final(PM)
  - Chart, presentation, assignment(PCA)
  - Comparative analysis of the result

V. PROPOSED SYSTEM

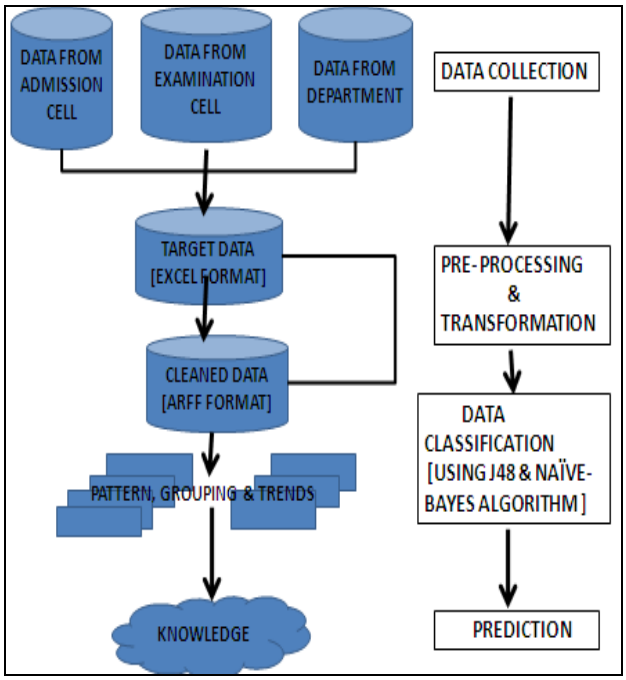

Figure 2 Proposed Systems

NAÏVE BAYES ALGORITHM:

The Naïve Bayes algorithm is a simple probabilistic classifier which is used on Bayes theorem with independence assumptions. It is one of the most basic classification techniques with various application spam detection, personal email sorting, document categorization and sentiment detection. Despite the Naïve design and over simplified assumptions this Naïve Bayes algorithm performs well in many complex real-world problems.

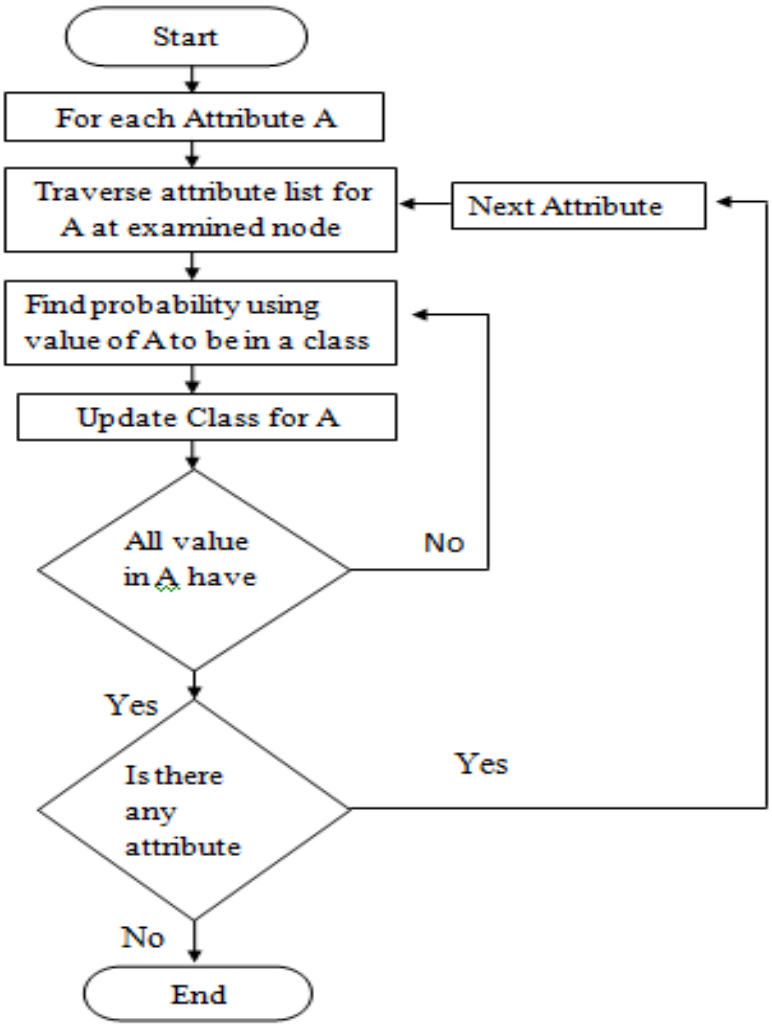

Figure3 Flow chart of Naive Bayes

J48 Algorithm

The j48 algorithm is a extension of ID3 algorithm. ID3 is an algorithm invented by Ross Quinlan used to generate a decision tree from the dataset. ID3 is typically used in the machine learning and natural language processing domains

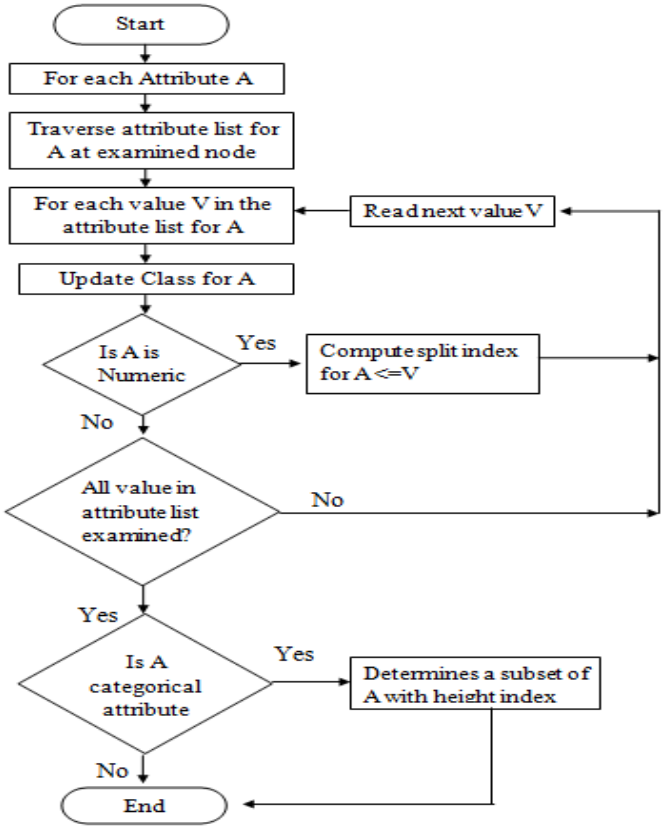

Figure 4 Flowchart of J48 algorithms:

VI. RESULT AND ANALYSIS

Implementation of AP Using J48:

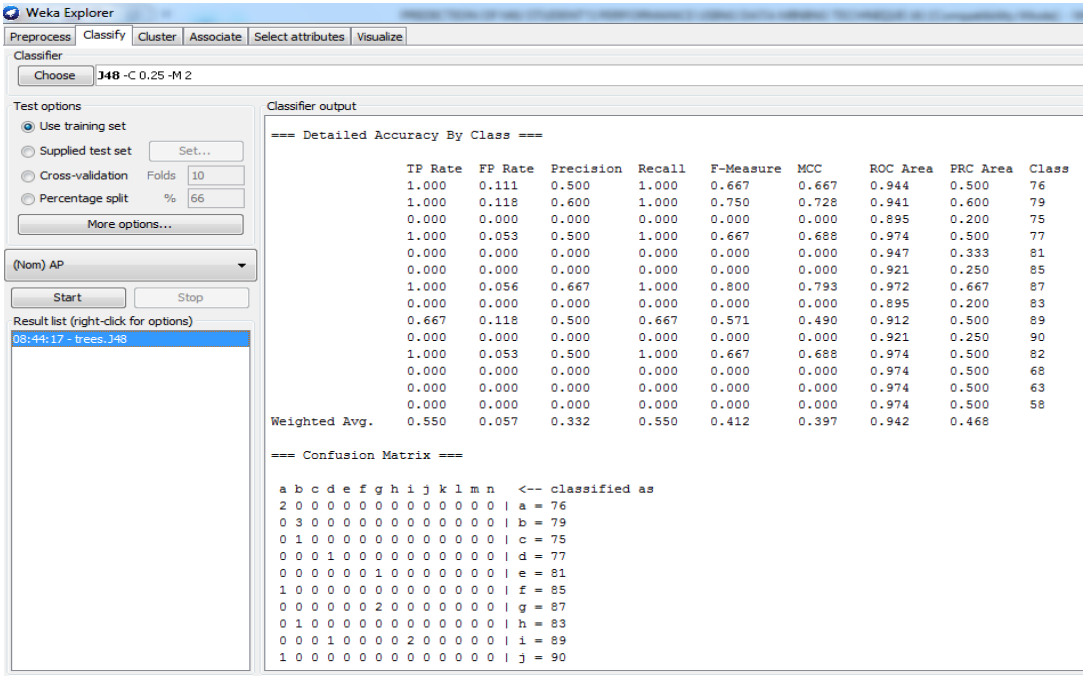

Figure 5: Implementation AP using J48

Implemantation Tree View of Decision Tree Based on AP using J48

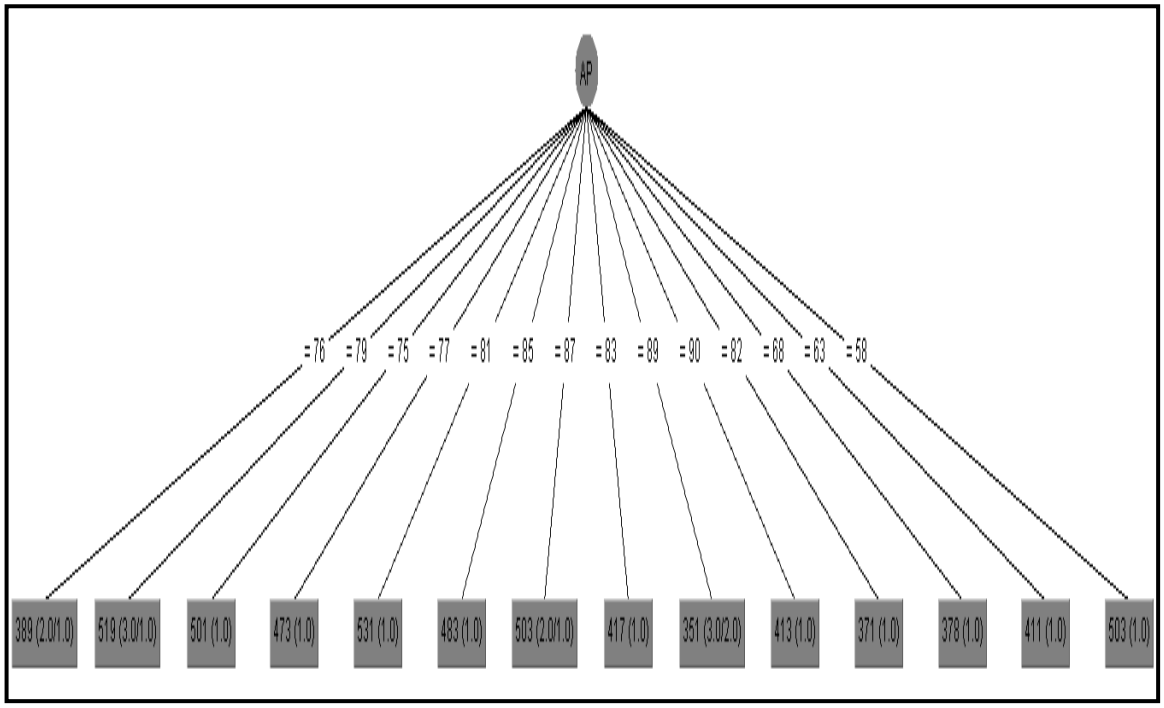

Figure 6: Tree view Implementation AP using J48

Implementation of ACP Using J48:

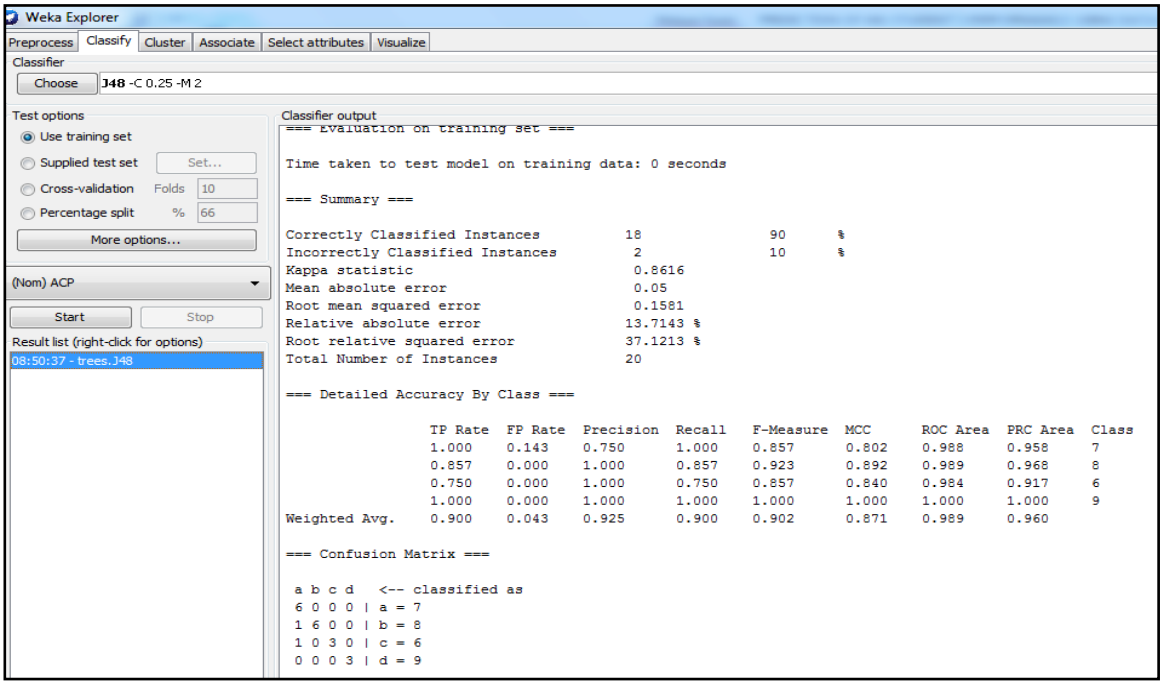

Figure 7: Implementation ACP using J48

Implemantation Tree View of Decision Tree Based on ACP using J48

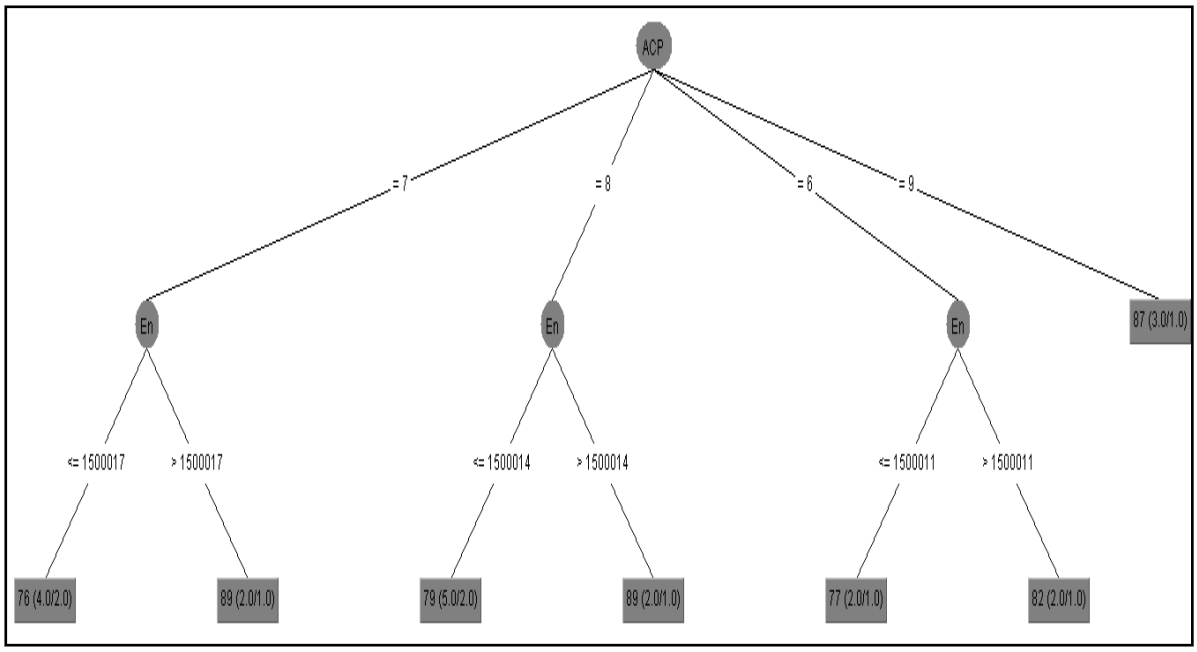

Figure 8: Tree View Implementation ACP Using J48

Implementation of PM using J48

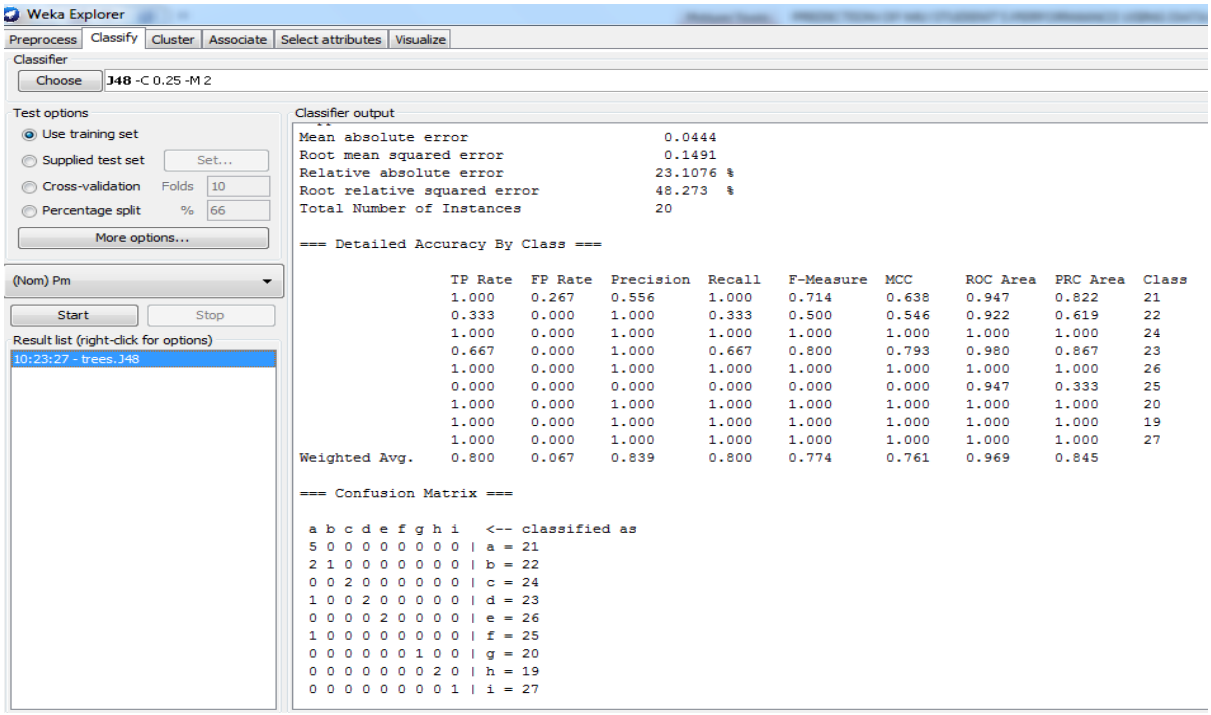

Figure 9: Implementation PM using J48

Implemantation Tree View of Decision Tree Based on PM using J48

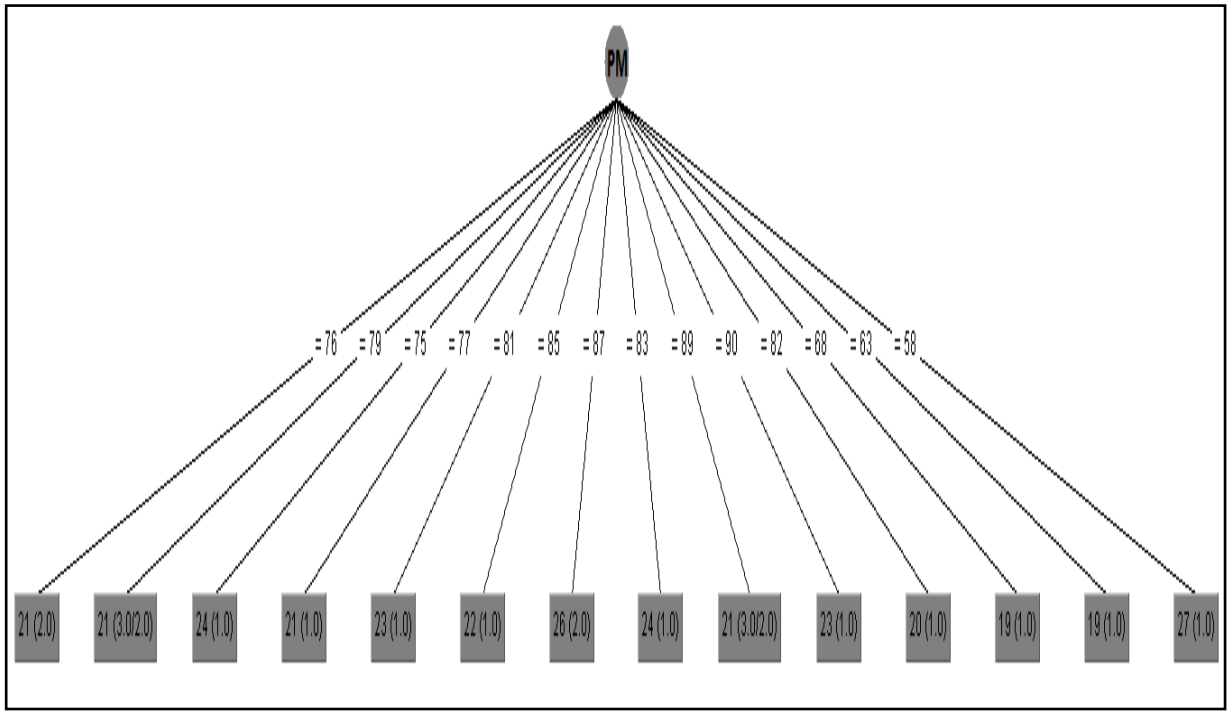

Figure10: Tree View Implementation PM using J48

Implementation of AP using Naïve Bayes Algorithm

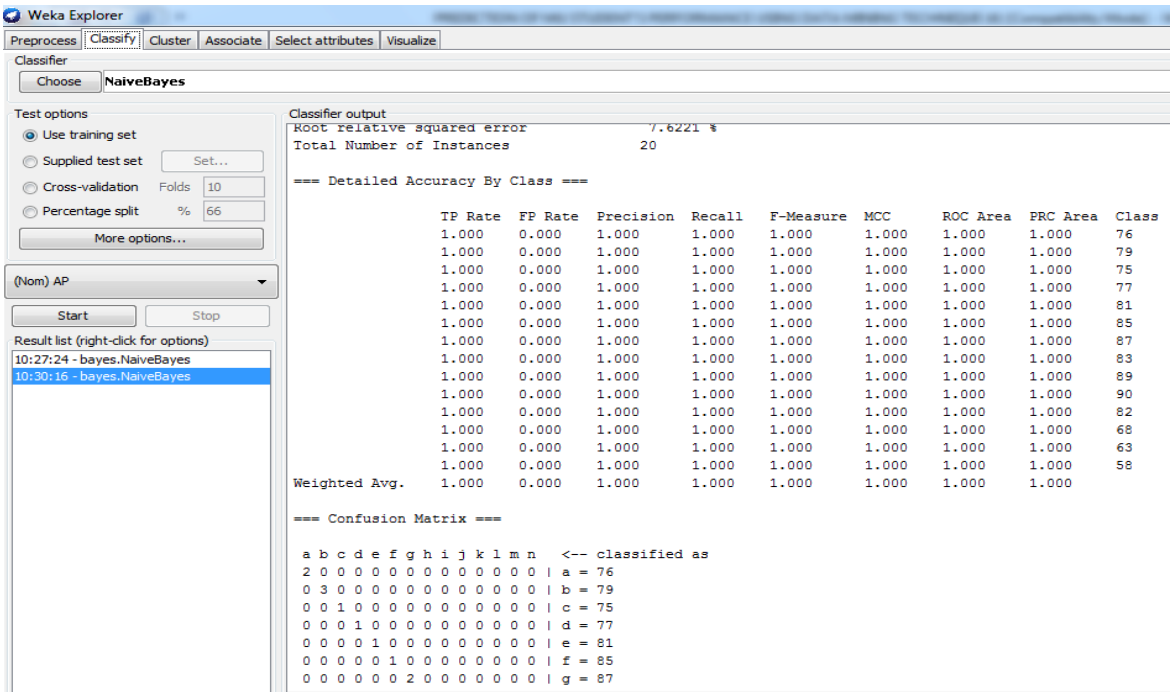

Figure 11: Naïve Bayes AP implementation

Implementation of ACP using Naïve Bayes Algorithm

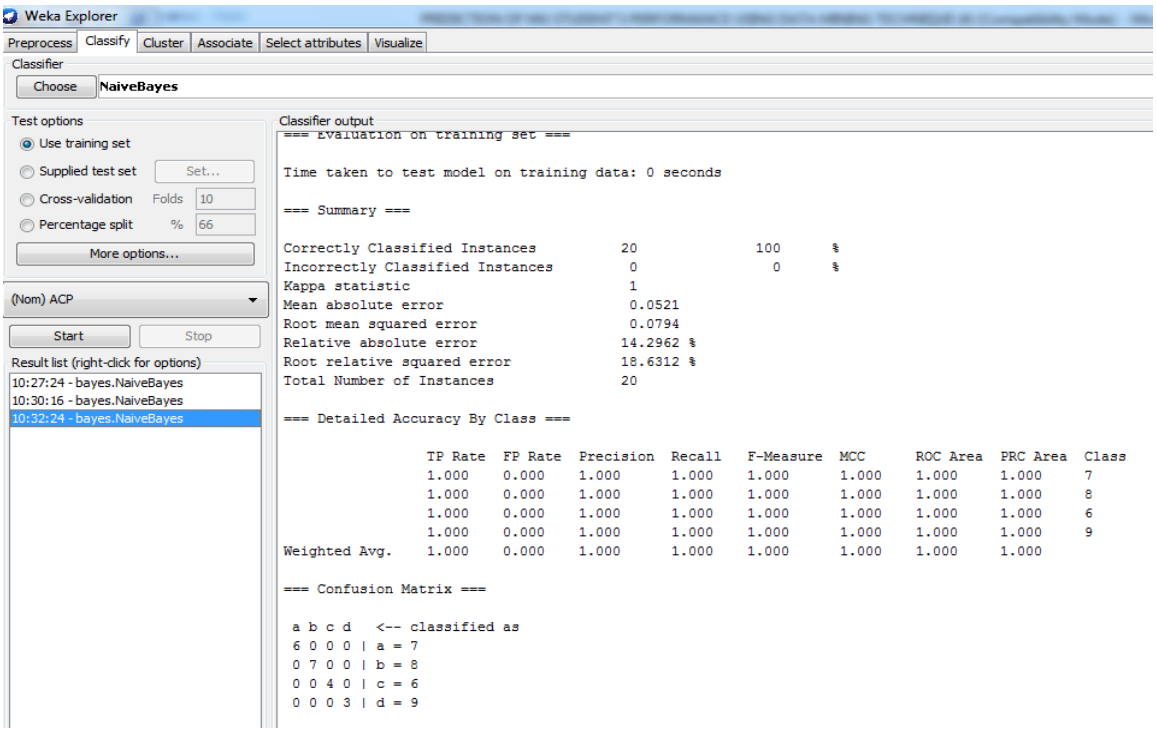

Figure 12: Naïve Bayes ACP implementation

Implementation of PM using Naïve Bayes Algorithm

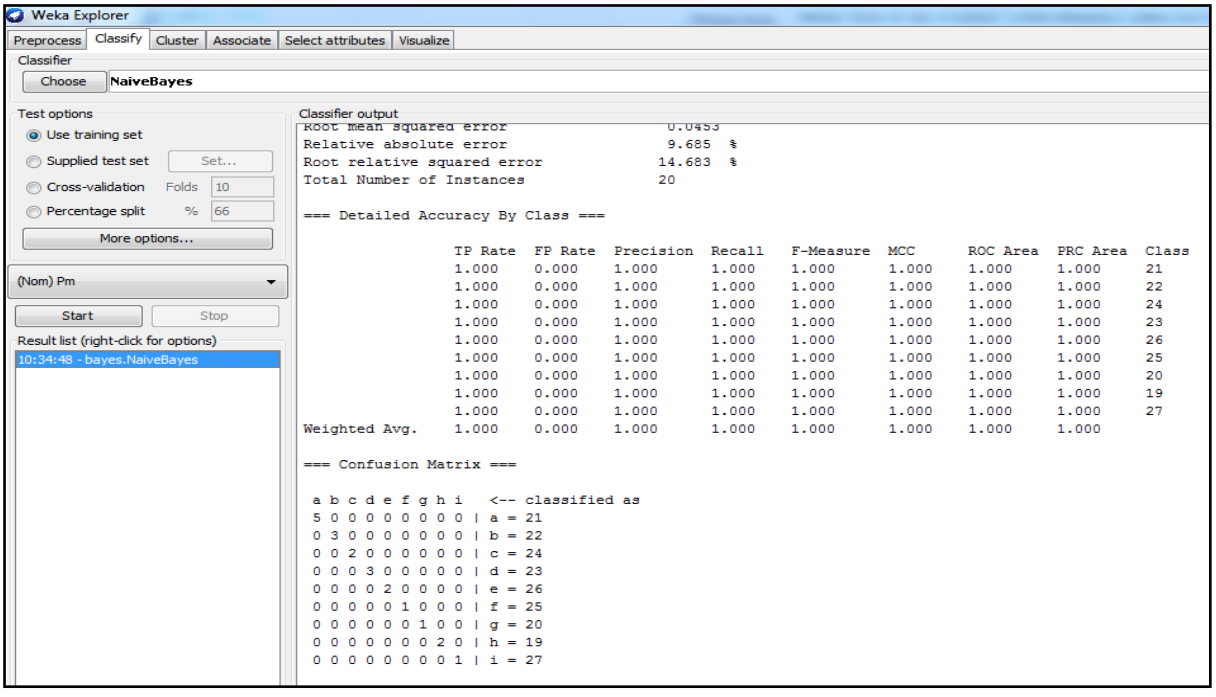

Figure 13: Naïve Bayes PM Implementation

Result Analysis of 20 Instances on the basis of confusion matrix

Method	Naive Bayes	J48
PM	100%	75%
AP	00%	25 %
ACP	0.0004	0.0296
TP	75%	100%
TP+2	90%	100%
TA	83.2341	100%
Total Number of Instances	20	20

Table 1: Result analysis of 20 Instances

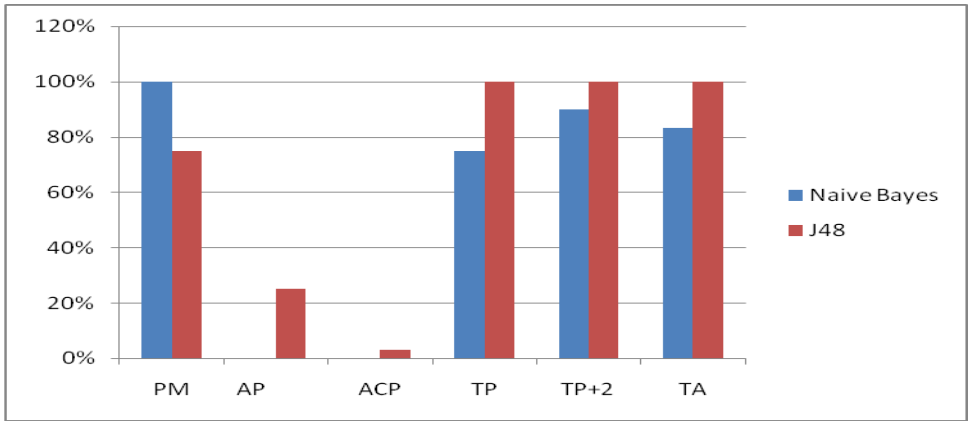

Figure 14: Graph representation of 20 instance result

Result Analysis of 153 Instances on the basis of confusion matrix

Method	Naive Bayes	J48
PM	56.8627 %	96.0784 %
AP	43.1373 %	3.9216 %
ACP	0.563	0.9603
TP	65%	100%
TP+2	80%	90%
TA	70%	95
Total Number of Instances	153	153

Table 2: Result analysis of 153 Instances

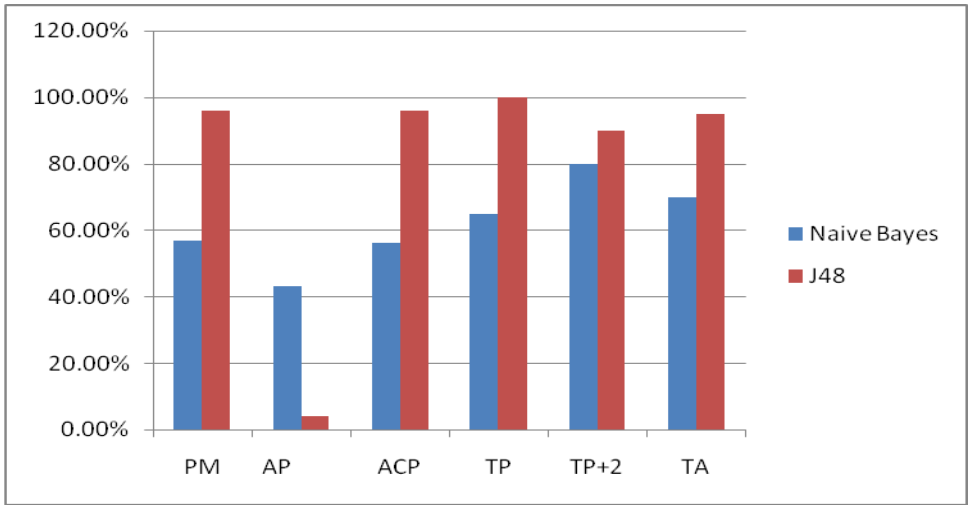

Figure 15: Graph representation of 153 instance result

VII. CONCLUSIONS

As number of records increases the accuracy of result cases very slowly for all the attributes in Naïve Bayes as per experiment result. While in case of J48 algorithm accuracy rate also decreases for only one attribute that is prefinal marks (PM). For other two attributes accuracy increases. There are huge difference between J48 and naïve Bayes results. In percentage this is 40% for 153 records. For small numbers of records naïve bayes classification algorithm is best because it show more accuracy as per experimental results. For large number of records J48 algorithm is better than naïve bayes algorithm because it shows more accuracy per instances.

VIII. FUTURE SCOPE

In future, accuracy rate can be calculated on actual data for different – different organizations by modifying or changing attributes. Anyone can apply different – different methods to know the best method suited for education domain.

ACKNOWLEDGEMENT

Foremost, I would like to express my sincere gratitude to my advisor Shiv Kumar for the continuous support of my study, for his patience, motivation, enthusiasm, and immense knowledge. Besides my advisor, I would like to thank the rest of my Friends and well-wisher.

REFERENCES

[1] Students Programming Skills", International Journal on Computer Science and Engineering Vol. 02, No. 03, pp 687-690.

[2] Zaïane, O. (2001)," Web usage mining for a better web-based learning environment", Proceedings Of Conference L.Arockiam, S.Charles,Arulkumar et.al(2010), "Deriving Association between Urban and Rural on Advanced Technology For Education, 60-64.

[3] Zaïane, O. (2002), "Building a recommender agent for e-learning systems". Proceedings of the International Conference on Computers in Education,55–59.

[4] Baker, R.S., Corbett, A.T., Koedinger, K.R. (2004), "Detecting Student Misuse of Intelligent Tutoring Systems". Proceedings of the 7th International Conference on Intelligent Tutoring Systems, 531-540.

[5] Tang, T., McCalla, G. (2005)," Smart recommendation for an evolving e-learning system: architecture and experiment", International Journal on E-Learning, vol. 4, issue1, 105–129.

[6] Merceron, A., Yacef, K. (2003)," A web-based tutoring tool with mining facilities to improve learning and teaching". Proceedings of the 11th International Conference on Artificial Intelligence in Education,

[7] M.Ramaswami and R.Bhaskaran(2010), "A CHAID Based Performance Prediction Model in Educational Data Mining", International Journal of Computer Science Issues Vol. 7, Issue 1, pp 10-18.
[8] Dringus, L.P., Ellis, T. (2005)," Using data mining as a strategy for assessing asynchronous discussion forums", Computer and Education Journal, 45, 141–160.
[9] M.Ramaswami and R.Bhaskaran(2010), "A CHAID Based Performance Prediction Model in Educational Data Mining", International Journal of Computer Science Issues Vol. 7, Issue 1, pp 10-18.
[10] Nguyen Thai-Nghe, Andre Busche, and Lars Schmidt-Thieme(2009), "Improving Academic Performance Prediction by Dealing with Class Imbalance", Ninth International Conference on Intelligent Systems Design and Applications, L.Arockiam, S.Charles, Arul Kumar et.al(2010), "Deriving Association between Urban and Rural Students Programming Skills", International Journal on Computer Science and Engineering Vol. 02, No. 03, pp 687 -690

