International Journal of Electronics Engineering (ISSN: 0973-7383)
Volume 11 • Issue 1 pp. 773-779 Jan 2019-June 2019 www.csjournals.com

Comparative Analysis of Naive Bayes and J48 Algorithms Using Data Mining Technique

Ankit Kumar Navalakha[1], Shrawan Kumar Sharma[2]

[1]Asst. Professor, Department of Computer Science, Mewar University, Gangrar, Chittorgarh, 312001, India
[1]Asst. Professor, Department of Computer Science, Vision Group of Colleges, Chittorgarh, 312001, India

Abstract: Data Mining is a computational technique or process of discovering patterns in large data sets and values involving the machine learning, mathematical, Statistics, and database system. It performs the classification of data using algorithms techniques. In this paper we use two data mining algorithms Navie Bayes and J48.we can perform the some data set of 10th, 12th marks, last semester, present semester, and student attendance data set record using data mining classification technique. We can compare the both algorithms on the basis of those data set records and find the best classification algorithms. We can also be finding the predication of those data set records.
Keywords: Data mining, Prediction, Naive ayes, J48, Machine Learning, Classification.

I. INTRODUCTION

Data mining is a process used by companies to turn raw data into useful information. By using software to look for patterns in large batches of data, businesses can learn more about their customers and develop more effective marketing strategies as well as increase sales and decrease costs. Data mining depends on effective data collection and warehousing as well as computer processing[1]
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:

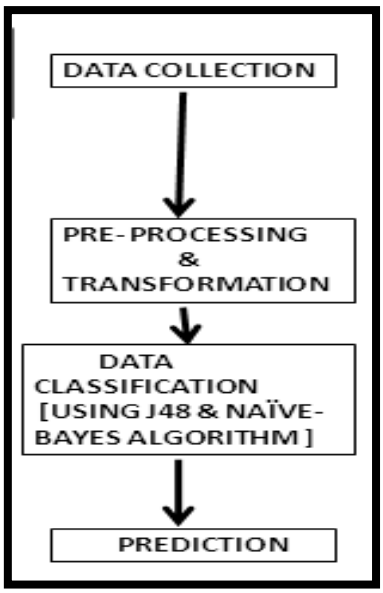


Figure 1: Data mining steps

II. PROBLEM STATEMENT

In current system the predicting of student performance is very difficult and more challenging due to the large data set off of educational database. In any educational database system facing two problem. First, existing prediction techniques is insufficient to find the most suitable methods or techniques to identify the most predicting the performance of students. Seconds the lack of investigations on the factors affecting student's achievements in particular courses.

International Journal of Electronics Engineering (ISSN: 0973-7383)
Volume 11 • Issue 1 pp. 773-779 Jan 2019-June 2019 www.csjournals.com

III. OBJECTIVE

The main object of this research is as following:

- Study of Data mining techniques
- Prediction result using Naïve Bayes & J48 algorithm
- Implementation of these algorithms on Weka tool following parameter:

  - Tenth marks
  - Intermediate marks
  - Attendance
  - Internal marks
  - Comparative analysis of the result

IV. PROPOSED SYSTEM:

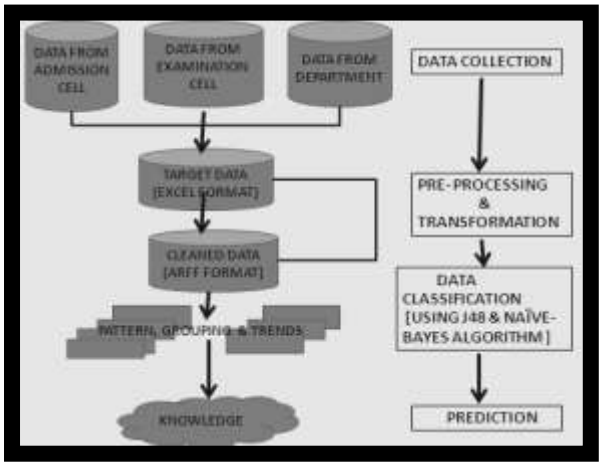
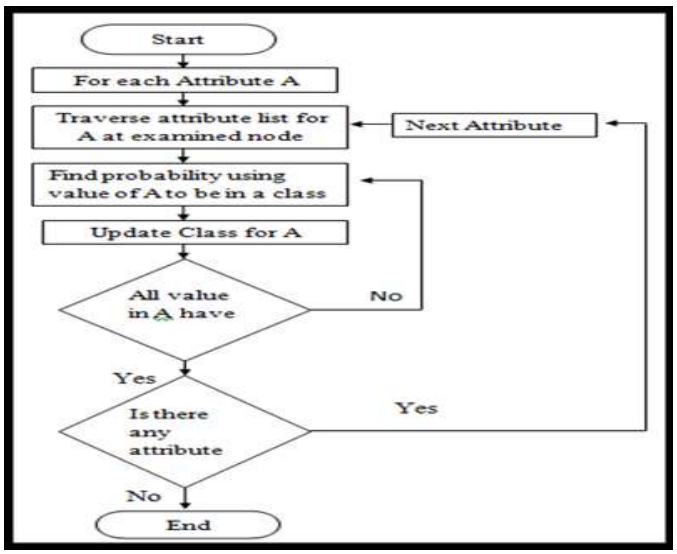


Figure 2: Proposed Systems

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.



International Journal of Electronics Engineering (ISSN: 0973-7383)
Volume 11 • Issue 1 pp. 773-779 Jan 2019-June 2019 www.csjournals.com

Figure 3: 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

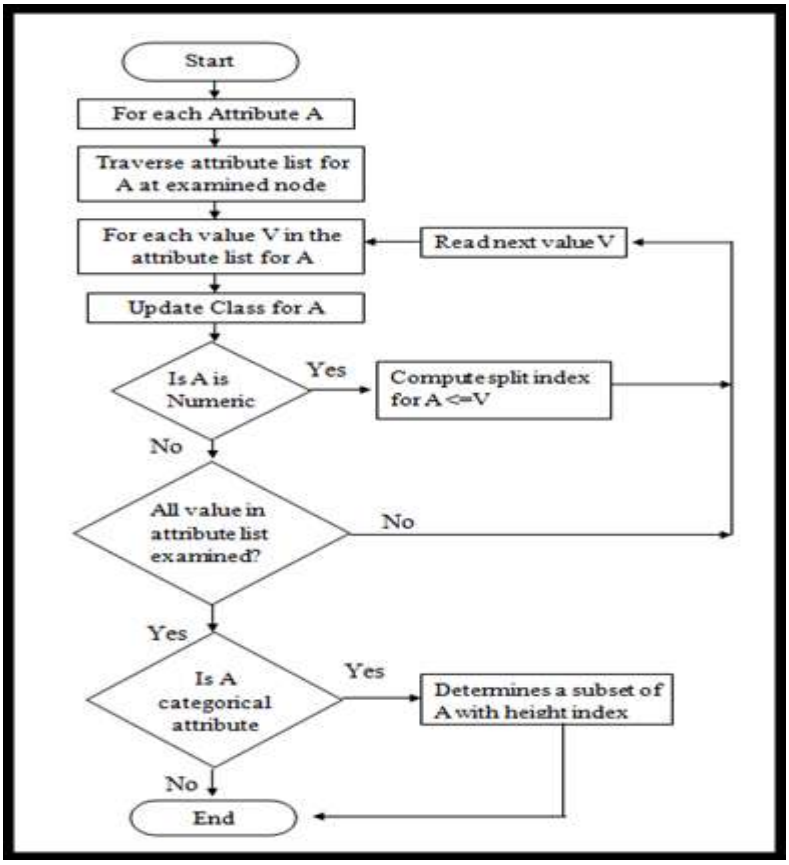


Figure 4: Flowchart of J48 algorithms

V. RESULT AND ANALYSIS

Implementation of Attendance Using J48:

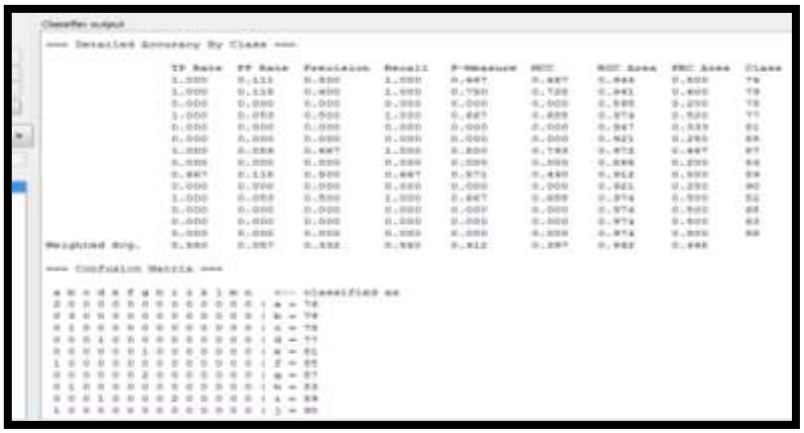


Figure 5: Implementation Attendance using J48

International Journal of Electronics Engineering (ISSN: 0973-7383)
Volume 11 • Issue 1 pp. 773-779 Jan 2019-June 2019 www.csjournals.com

Implemantation Tree View of Decision Tree Based on Attendance using J48

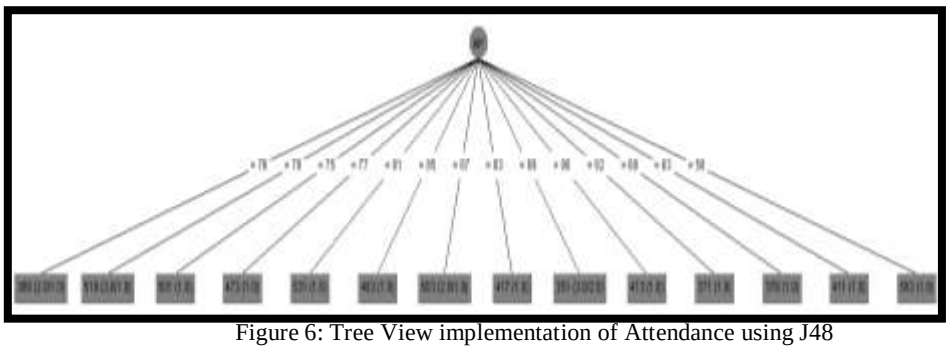
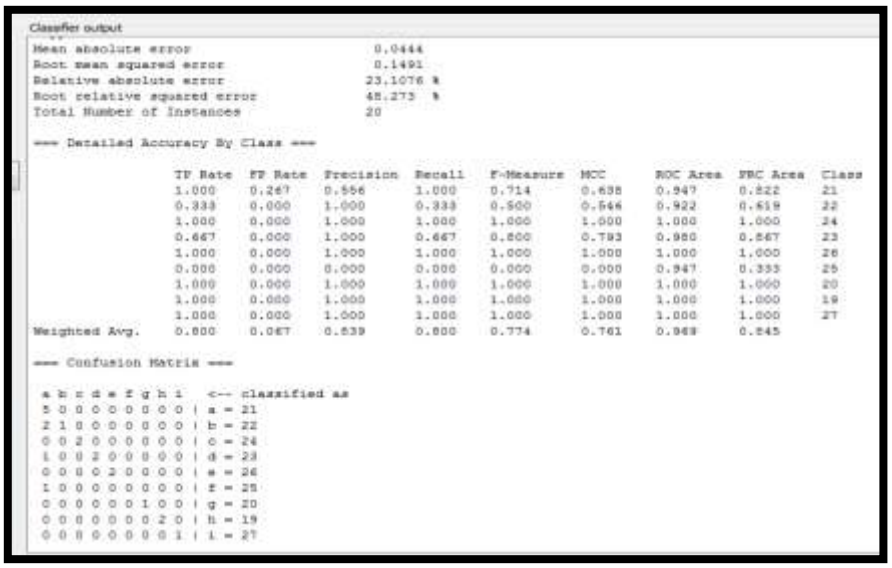


Figure 6: Tree View implementation of Attendance using J48

Implementation of Internal Marks using J48



```
Classifier output

Mean absolute error                      0.0444
Root mean squared error                  0.1491
Relative absolute error                 23.1076 %
Root relative squared error             45.273  %
Total Number of Instances               20

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
                 1.000    0.267    0.556      1.000   0.714      0.638    0.947     0.822     21
                 0.333    0.000    1.000      0.333   0.500      0.546    0.922     0.619     22
                 1.000    0.000    1.000      1.000   1.000      1.000    1.000     1.000     24
                 0.667    0.000    1.000      0.667   0.800      0.793    0.980     0.867     23
                 1.000    0.000    1.000      1.000   1.000      1.000    1.000     1.000     26
                 0.000    0.000    0.000      0.000   0.000      0.000    0.947     0.333     25
                 1.000    0.000    1.000      1.000   1.000      1.000    1.000     1.000     20
                 1.000    0.000    1.000      1.000   1.000      1.000    1.000     1.000     19
                 1.000    0.000    1.000      1.000   1.000      1.000    1.000     1.000     27
Weighted Avg.    0.800    0.067    0.839      0.800   0.774      0.761    0.969     0.845

=== Confusion Matrix ===

 a b c d e f g h i   <-- classified as
 5 0 0 0 0 0 0 0 0 | a = 21
 2 1 0 0 0 0 0 0 0 | b = 22
 0 0 2 0 0 0 0 0 0 | c = 24
 1 0 0 2 0 0 0 0 0 | d = 23
 0 0 0 0 2 0 0 0 0 | e = 26
 1 0 0 0 0 0 0 0 0 | f = 25
 0 0 0 0 0 0 1 0 0 | g = 20
 0 0 0 0 0 0 0 2 0 | h = 19
 0 0 0 0 0 0 0 0 1 | i = 27
```

Figure 7: Implementation internal marks using J48

Implemantation Tree View of Decision Tree Based on Internal Marks using J48

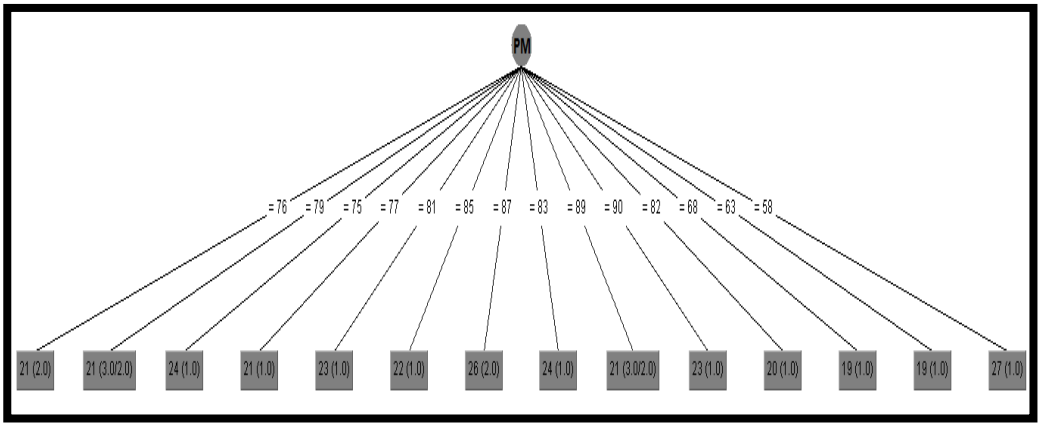


Figure 8: Tree View Implementation Internal Marks using J48

International Journal of Electronics Engineering (ISSN: 0973-7383)
Volume 11 • Issue 1 pp. 773-779 Jan 2019-June 2019 www.csjournals.com

Implementation of Attendance using Naïve Bayes Algorithm

Figure 9: Naïve Bayes Attendance implementation

Implementation of Internal Marks using Naïve Bayes Algorithm

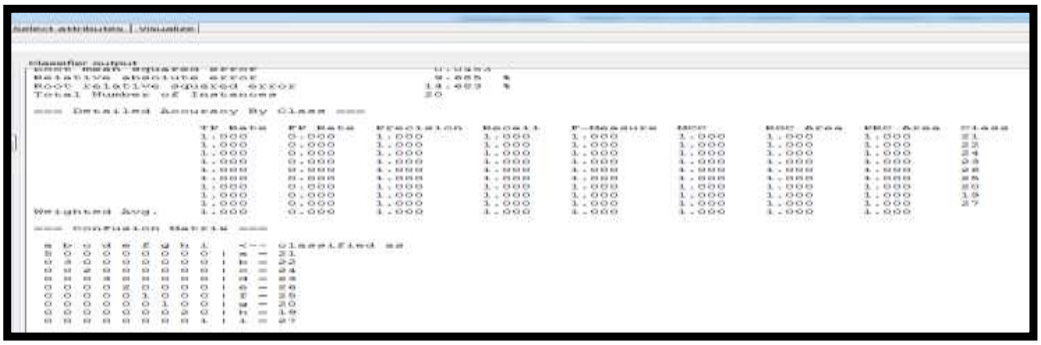


Figure 10: Naïve Bayes Internal Marks Implementation

Result Analysis of 20 Instances on the basis of confusion matrix

Table 1: Result analysis of 50 Instances

Method	Naive Bayes	J48
Internal Marks	76%	90%
Attendance	10%	25 %
Tenth marks	75%	100%
Intermediate Marks	90%	100%
Total Number of Instances	50	50

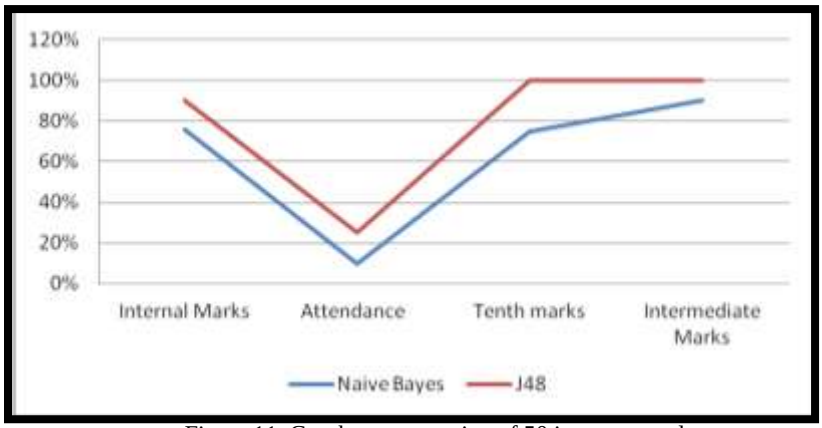


Figure 11: Graph representation of 50 instance result

International Journal of Electronics Engineering (ISSN: 0973-7383)
Volume 11 • Issue 1 pp. 773-779 Jan 2019-June 2019 www.csjournals.com

Result Analysis of 153 Instances on the basis of confusion matrix

Table 2: Result analysis of 120 Instances

Method	Naive Bayes	J48
Internal Marks	46.8627 %	86.0784 %
Attendance	33.1373 %	3.9216 %
Tenth marks	45%	80%
Intermediate Marks	60%	70%
Total Number of Instances	120	120

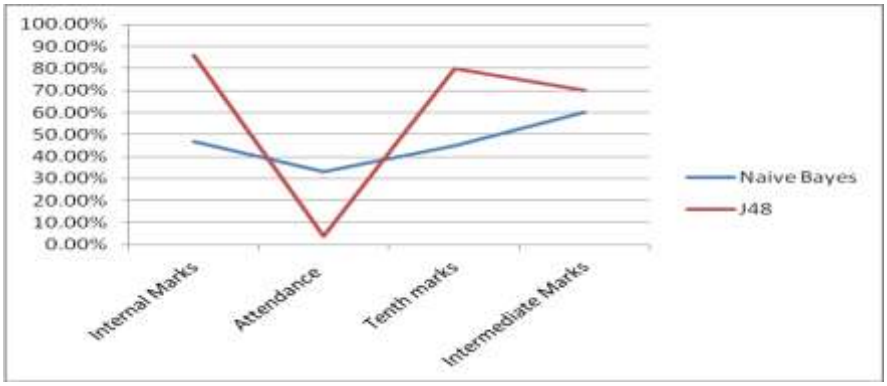


Figure 12: Graph representation of 120 instance result

VI. CONCLUSIONS

As our predication the data set records the both algorithms they have huge difference between the algorithms. The J48 is performing better as per navie bayer. The J48 perform much better accuracy as experimental result. In small number (50 data set) and the large data set (120 data set) both these set the J48 is perform better.

ACKNOWLEDGEMENT

Foremost, I would like to express my sincere gratitude to my advisor Mr..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.