Associative Rule Mining in Student-centric Assessment and Evaluation

Dr.K.Kasturi[1] and C.Sreedevi[2]

[1]*Assistant Professor, Department of Information Technology,
Vels Institute of Science, Technology & Advanced Studies (VISTAS), Chennai,
E-mail: kasturi.scs@velsuniv.ac.in*

[2] *Research Scholar, Department of Computer science
Vels Institute of Science, Technology & Advanced Studies (VISTAS), Chennai,
E-mail: sree_gokul@hotmail.com*

Abstract

Education is imparted by the established academic institutions .Nowadays the big challenge of school students are getting through the competitive examinations and it also shows the quality of our education. By the assessment and evaluation of the student‘s ability the quality of the education inculcated can be improved. This paper deals with assessments that are student centric to enhance their ability that paves the way for their career development. Data mining techniques such as association rule mining have been applied to the data to analyse and find the factors that are really needed to assess the student capability. By applying associative rule mining the association between the assessment factors that are more significantly affecting the performance of a student has been identified and it provides a decision support for teaching and the assessment curriculum.

Keywords: *Data Mining, Associative Rule Mining, Apriori algorithm, student-centric assessment.*

1. Introduction

Commonly assessment involves the student with addressing written questions, perhaps under a time limit, although more recently much attention was given to the use of oral assessment. Student performance is an essential justification for assessing the quality of teaching and the significant sign that students are mastering the knowledge they have learnt. Exams like Olympiad are much more conceptual and tricky, helping students grasp the subject well. In the end, this increases the performance of regular classes. These examinations motivate to develop in children a kind of analytical thinking that is useful in any examination, whether it is IIT, NEET or any other competitive examination. Such tests not only measure the basic concepts taught in the classroom, they also improve a child's ability to think analytically. This improves the ability to think, problem-solve skills, trust and thus helps in a young child's overall growth. This paper deals with the unique assessment techniques for students to get through their competitive examinations and the prediction have been justified by the data mining techniques.

2. Associative Rule Mining

Associative Rule Mining often known as ARM is the study of knowing a concept by association. They are searching for interconnections between the factors as the variables and findings as the outcomes. Association rule mining is a method of data mining to discover interesting relationships among features. “Association rule was applied in admission data to find some knowledge for supporting admission planning” [1]. “The Machine Learning approach can be objectively deployed to obtain a predictive relationship and behavioural aspects that permits mapping the interaction behaviour of students with their course outcome using associative rule mining for feature selection techniques”[2]. “A framework for an effective educational process using data-mining techniques to uncover the hidden trends and patterns and making accuracy based predictions through a higher level of analytical sophistication in the process of counselling students"[3]. The Mining of association has two essential measures. First phase, the generation of frequent itemsets, is a process to identify frequent itemsets that have a minimum support and a minimum confidence value greater than the assigned threshold values. Second phase, the generation of rules, is a process of generating rules from frequent itemset. Association parameters-support, confidence, and lift values evaluate the rules produced. Most probably many authors use apriori algorithm for associative rule mining [10].The major advantage over this algorithm is that it creates strong rules and pruned rules (repeated or redundant rules are removed to reduce the computational time).Nowadays for decision making and prediction data mining techniques have been used in many of the android applications also.[8]

```
INPUT: S, support where S = dataset, min_support = real
OUTPUT: set of Frequent Itemsets
Require : S ≠ Ø ,0 ≤ min_support ≤ 1
procedure GetFrequentItemsets
    freqSets[ ] ← null
    for all Itemsets i in S do
        if support ≥ min_support then
            freqSets[ ] ← i
        end if
    end for
end procedure
```

Table 1: Apriori Algorithm

3. Student Assessment

Student master any material and this is more than memorizing the information. The students are able to use the knowledge to reason after information understanding and to address other types of problems. The potential of students to illustrate such skills or practices, such as group events, middle school science lab, public speaking or presentations [4]. Another way for students to achieve academic success is by producing quality products that show in their content that the student has acquired practical knowledge, essential logic and problem-solving skills and applicable production skills, such as a science model, report or project. A series of questions are posed to the respondent, each followed by a selection of alternative answers. The role of the respondent is to pick the right answer, or the best response. This may include several options, true / false, match, quick reply, fill-in objects. [4]

S.No.	Half-yearly_exam	ASS_1	ASS_2	ASS_3	ASS_4	Competitve_exams
1	good	vgood	vgood	vgood	good	Yes
2	average	vgood	good	vgood	good	Yes
3	vgood	good	good	good	vgood	Yes
4	good	average	vgood	good	average	Yes
5	good	average	good	average	average	No
6	good	good	average	good	poor	No
7	good	good	good	good	vgood	Yes
8	vgood	good	vgood	good	good	Yes
9	good	average	good	poor	poor	No
10	average	good	average	good	average	No

Table 2: Sample Data

ISSN: 2005-4238 IJAST

The data has been collected from 8th and 9th standard school students to evaluate their assessment and to improve their academic knowledge more applicable and get through into the national competitive exams like Olympiad, National Talent Search Examination (NTSE) or KVPY. The assessment parameters are as follows:

1. **Content Assessment** as Content mastering – **Half –yearly Exam**
2. **Reasoning Assessment** as analysing and problem solving - **ASS_1**
3. **Skill Assessment** as group activities-**ASS_2**
4. **Product Assessment** as science model/report/project- **ASS_3**
5. **Response Assessment** as Multiple Choice Questions-**ASS_4**

All the five assessment parameters have four levels of values: very good, good, average and poor. The range of values for vgood (86% - 100%), good (71%-85%), average (55%-70%), poor(55%-40%).The outcome or the target parameter: **Competitive_exams.** The target variable has two values yes/no implies that a student is able/not able to clear the exams.

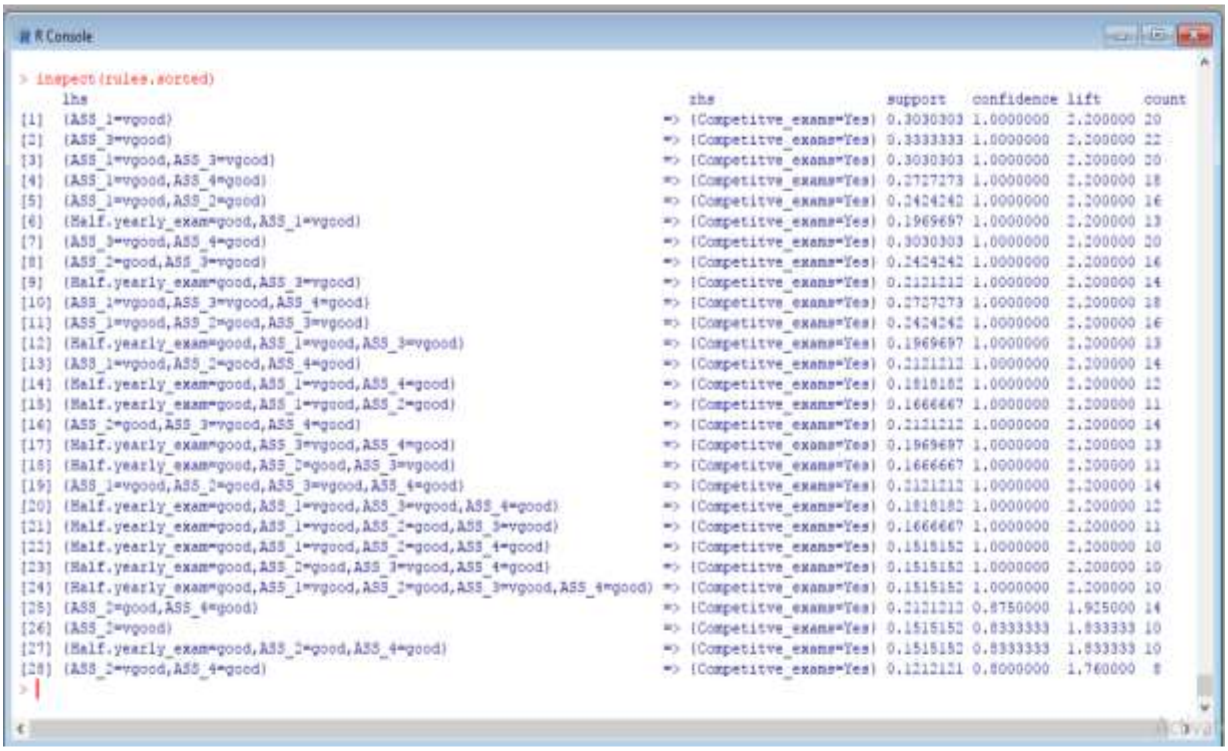
4. Mining Rules Using R

R is a popular statistical data mining tool and the apriori algorithm has been implemented in R programming for the student dataset.

Fig.1: Rule mining ‘Competitive_exams=Yes’

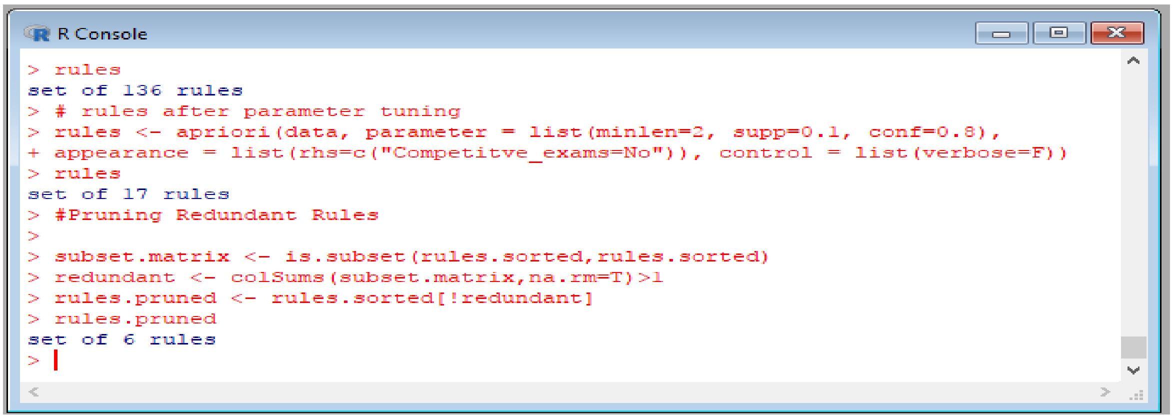
In Fig.1 by using apriori as default 136 rules have been got from the dataset and by assigning the values for support and confidence the rules based on the frequently occurring item sets the rules drawn are 28 rules for the target variable with “Yes” value. Strong association rules have support and confidence greater than or equal to a user-specified minimum support threshold and respectively a minimum confidence threshold and draws 4 rules as strong rules.

In Fig.2 LHS shows the five assessment parameters and the RHS shows the target variable competitive_exams as “Yes”. The figure shows 28 inspected rules for the students who could clear the competitive exams.



```
R Console
> inspect(rules.sorted)
     lhs                                                                   rhs                        support   confidence lift     count
[1]  {ASS_1=vgood}                                                      => {Competitve_exams=Yes} 0.3030303 1.0000000  2.200000 20
[2]  {ASS_3=vgood}                                                      => {Competitve_exams=Yes} 0.3333333 1.0000000  2.200000 22
[3]  {ASS_1=vgood,ASS_3=vgood}                                          => {Competitve_exams=Yes} 0.3030303 1.0000000  2.200000 20
[4]  {ASS_1=vgood,ASS_4=good}                                           => {Competitve_exams=Yes} 0.2727273 1.0000000  2.200000 18
[5]  {ASS_1=vgood,ASS_2=good}                                           => {Competitve_exams=Yes} 0.2424242 1.0000000  2.200000 16
[6]  {Half.yearly_exam=good,ASS_1=vgood}                                => {Competitve_exams=Yes} 0.1969697 1.0000000  2.200000 13
[7]  {ASS_3=vgood,ASS_4=good}                                           => {Competitve_exams=Yes} 0.3030303 1.0000000  2.200000 20
[8]  {ASS_2=good,ASS_3=vgood}                                           => {Competitve_exams=Yes} 0.2424242 1.0000000  2.200000 16
[9]  {Half.yearly_exam=good,ASS_3=vgood}                                => {Competitve_exams=Yes} 0.2121212 1.0000000  2.200000 14
[10] {ASS_1=vgood,ASS_3=vgood,ASS_4=good}                               => {Competitve_exams=Yes} 0.2727273 1.0000000  2.200000 18
[11] {ASS_1=vgood,ASS_2=good,ASS_3=vgood}                               => {Competitve_exams=Yes} 0.2424242 1.0000000  2.200000 16
[12] {Half.yearly_exam=good,ASS_1=vgood,ASS_3=vgood}                    => {Competitve_exams=Yes} 0.1969697 1.0000000  2.200000 13
[13] {ASS_1=vgood,ASS_2=good,ASS_4=good}                                => {Competitve_exams=Yes} 0.2121212 1.0000000  2.200000 14
[14] {Half.yearly_exam=good,ASS_1=vgood,ASS_4=good}                     => {Competitve_exams=Yes} 0.1818182 1.0000000  2.200000 12
[15] {Half.yearly_exam=good,ASS_1=vgood,ASS_2=good}                     => {Competitve_exams=Yes} 0.1666667 1.0000000  2.200000 11
[16] {ASS_2=good,ASS_3=vgood,ASS_4=good}                                => {Competitve_exams=Yes} 0.2121212 1.0000000  2.200000 14
[17] {Half.yearly_exam=good,ASS_3=vgood,ASS_4=good}                     => {Competitve_exams=Yes} 0.1969697 1.0000000  2.200000 13
[18] {Half.yearly_exam=good,ASS_2=good,ASS_3=vgood}                     => {Competitve_exams=Yes} 0.1666667 1.0000000  2.200000 11
[19] {ASS_1=vgood,ASS_2=good,ASS_3=vgood,ASS_4=good}                    => {Competitve_exams=Yes} 0.2121212 1.0000000  2.200000 14
[20] {Half.yearly_exam=good,ASS_1=vgood,ASS_3=vgood,ASS_4=good}         => {Competitve_exams=Yes} 0.1818182 1.0000000  2.200000 12
[21] {Half.yearly_exam=good,ASS_1=vgood,ASS_2=good,ASS_3=vgood}         => {Competitve_exams=Yes} 0.1666667 1.0000000  2.200000 11
[22] {Half.yearly_exam=good,ASS_1=vgood,ASS_2=good,ASS_4=good}          => {Competitve_exams=Yes} 0.1515152 1.0000000  2.200000 10
[23] {Half.yearly_exam=good,ASS_2=good,ASS_3=vgood,ASS_4=good}          => {Competitve_exams=Yes} 0.1515152 1.0000000  2.200000 10
[24] {Half.yearly_exam=good,ASS_1=vgood,ASS_2=good,ASS_3=vgood,ASS_4=good} => {Competitve_exams=Yes} 0.1515152 1.0000000  2.200000 10
[25] {ASS_2=good,ASS_4=good}                                            => {Competitve_exams=Yes} 0.2121212 0.8750000  1.925000 14
[26] {ASS_2=vgood}                                                      => {Competitve_exams=Yes} 0.1515152 0.8333333  1.833333 10
[27] {Half.yearly_exam=good,ASS_2=good,ASS_4=good}                      => {Competitve_exams=Yes} 0.1515152 0.8333333  1.833333 10
[28] {ASS_2=vgood,ASS_4=good}                                           => {Competitve_exams=Yes} 0.1212121 0.8000000  1.760000  8
>
```

Fig.2: Inspected Rules ‘Competitive_exams=Yes’



```
R Console
> rules
set of 136 rules
> # rules after parameter tuning
> rules <- apriori(data, parameter = list(minlen=2, supp=0.1, conf=0.8),
+ appearance = list(rhs=c("Competitve_exams=No")), control = list(verbose=F))
> rules
set of 17 rules
> #Pruning Redundant Rules
>
> subset.matrix <- is.subset(rules.sorted,rules.sorted)
> redundant <- colSums(subset.matrix,na.rm=T)>1
> rules.pruned <- rules.sorted[!redundant]
> rules.pruned
set of 6 rules
>
```

Fig.3: Rule mining ‘Competitive_exams=No’

Here by using apriori as default 136 rules have been got from the dataset and by parameter specification for support and confidence the rules drawn are 17 rules for the target variable with “No” value and drawn 6 rules as strong rules.

```
R Console
> inspect(rules.sorted)
     lhs                                         rhs                          support   confidence lift     count
[1]  {ASS_1=poor}                             => {Competitve_exams=No} 0.1818182 1.0000000  1.941176 12
[2]  {ASS_4=poor}                             => {Competitve_exams=No} 0.1818182 1.0000000  1.941176 12
[3]  {ASS_3=average}                          => {Competitve_exams=No} 0.2424242 1.0000000  1.941176 16
[4]  {ASS_2=average}                          => {Competitve_exams=No} 0.2727273 1.0000000  1.941176 18
[5]  {Half.yearly_exam=good,ASS_1=poor}       => {Competitve_exams=No} 0.1212121 1.0000000  1.941176  8
[6]  {Half.yearly_exam=good,ASS_4=poor}       => {Competitve_exams=No} 0.1212121 1.0000000  1.941176  8
[7]  {ASS_2=average,ASS_3=average}            => {Competitve_exams=No} 0.1212121 1.0000000  1.941176  8
[8]  {ASS_3=average,ASS_4=average}            => {Competitve_exams=No} 0.1212121 1.0000000  1.941176  8
[9]  {Half.yearly_exam=good,ASS_3=average}    => {Competitve_exams=No} 0.1515152 1.0000000  1.941176 10
[10] {Half.yearly_exam=average,ASS_2=average} => {Competitve_exams=No} 0.1212121 1.0000000  1.941176  8
[11] {ASS_2=average,ASS_4=average}            => {Competitve_exams=No} 0.1515152 1.0000000  1.941176 10
[12] {ASS_1=good,ASS_2=average}               => {Competitve_exams=No} 0.1212121 1.0000000  1.941176  8
[13] {Half.yearly_exam=good,ASS_2=average}    => {Competitve_exams=No} 0.1515152 1.0000000  1.941176 10
[14] {ASS_4=average}                          => {Competitve_exams=No} 0.2424242 0.8888889  1.725490 16
[15] {ASS_1=average}                          => {Competitve_exams=No} 0.1818182 0.8571429  1.663866 12
[16] {Half.yearly_exam=good,ASS_1=average}    => {Competitve_exams=No} 0.1363636 0.8181818  1.588235  9
[17] {Half.yearly_exam=good,ASS_4=average}    => {Competitve_exams=No} 0.1363636 0.8181818  1.588235  9
>
```

Fig. 4: Inspected Rules 'Competitive_exams=No'

Here LHS shows the five assessment parameters and the RHS shows the target variable competitive_exams as "No". The figure shows the 17 rules for the students who could not clear the competitive exams.

4.1. Pruned Rules

The pruning of rules deals with removing the infrequent itemsets that are redundant(repetitive) which are unnecessary in finding the significant association.

```
R Console
> rules.pruned <- rules.sorted[!redundant]
> inspect(rules.pruned)
    lhs                        rhs                      support   confidence lift     count
[1] {ASS_1=vgood}           => {Competitve_exams=Yes} 0.3030303 1.0000000  2.200000 20
[2] {ASS_3=vgood}           => {Competitve_exams=Yes} 0.3333333 1.0000000  2.200000 22
[3] {ASS_2=good,ASS_4=good} => {Competitve_exams=Yes} 0.2121212 0.8750000  1.925000 14
[4] {ASS_2=vgood}           => {Competitve_exams=Yes} 0.1515152 0.8333333  1.833333 10
>
```

Fig 5: Pruned Rules 'Competitive_exams=Yes'

Here by analysing the strong rules for "Competitive_exams=Yes" the ASS_1, ASS_3, ASS_2 and ASS_4 are obtained as the significant assessment parameters respectively that are more associated for **clearing** the competitive exams.

```
R Console
> rules.pruned <- rules.sorted[!redundant]
> inspect(rules.pruned)
    lhs                rhs                   support   confidence lift     count
[1] {ASS_1=poor}    => {Competitve_exams=No} 0.1818182 1.0000000  1.941176 12
[2] {ASS_4=poor}    => {Competitve_exams=No} 0.1818182 1.0000000  1.941176 12
[3] {ASS_3=average} => {Competitve_exams=No} 0.2424242 1.0000000  1.941176 16
[4] {ASS_2=average} => {Competitve_exams=No} 0.2727273 1.0000000  1.941176 18
[5] {ASS_4=average} => {Competitve_exams=No} 0.2424242 0.8888889  1.725490 16
[6] {ASS_1=average} => {Competitve_exams=No} 0.1818182 0.8571429  1.663866 12
>
```

Fig 6: Pruned Rules 'Competitive_exams=No'

Here by analysing the strong rules for "Competitive_exams=No" the marks obtained by the students as 'poor' and 'average' for the parameters ASS_1, ASS_4, ASS_3 and ASS_2 shows that these assessment parameters are more associated for **Not clearing** the competitive exams.

4.2 Analysis of Association Rules

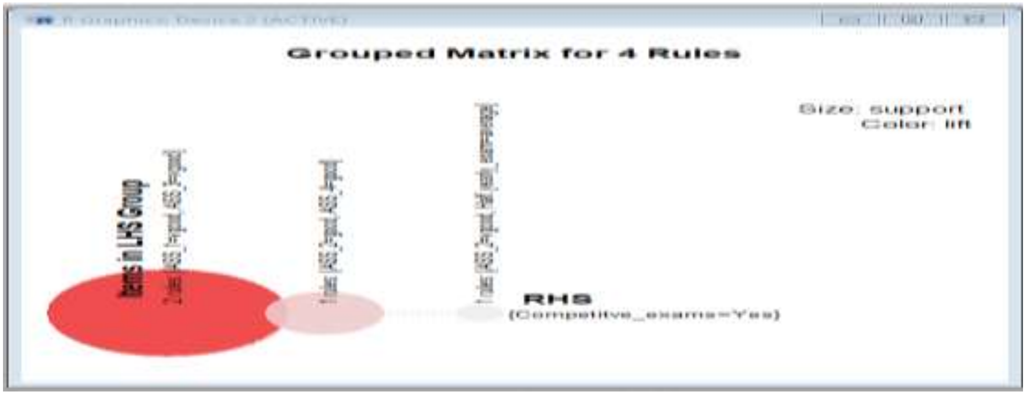


Fig 7: Grouped Matrix 'Competitive_exams=Yes'

Fig 8: Grouped Matrix 'Competitive_exams=No'

Support is the measure of how frequent an item set is in a dataset. Confidence is the conditional probability of occurrence of consequent (then) given the antecedent (if).
Lift is the ratio of Confidence to the Expected Confidence.
According to ARM using Apriori algorithm the association between the assessment factors and the outcome depicts that

- Firstly the students concentrating and scoring high in Reasoning and Product assessments have more chance of clearing Competitive Exams.
- Secondly the students concentrating and scoring high in Skill and Response assessments also have chance of clearing Competitive Exams.
- The students scoring average or poor scoring in Reasoning, product, skill and Response assessments not have the chance of clearing the Competitive Exams.

- The students scoring average or good marks in Half-yearly examination have chance of clearing or not clearing the Competitive Exams.

5. Conclusion

The pruned strong rules and the grouped matrix for "Yes", "No" values for the competitive exams shows that the assessment parameter 'Half-yearly_exam' does not have more impact on the target variable compared to other assessment parameters. From this inference the students can be more concentrated in the assessments like Reasoning ability, Project , Group Activity and MCQs etc.,. In order to compete the competitive world the students can be trained effectively by the teachers to make the concepts more applicable as the real world entities.

References

[1] Rojanavasu, P. (2019). Educational Data Analytics using Association Rule Mining and Classification. In 2019 Joint International Conference on Digital Arts, Media and Technology with ECTI Northern Section Conference on Electrical, Electronics, Computer and Telecommunications Engineering (ECTI DAMT-NCON) (pp. 142-145). IEEE.
[2] Singh, S., & Lal, S. P. (2013, November). Using feature selection and association rule mining to evaluate digital courseware. In 2013 Eleventh International Conference on ICT and Knowledge Engineering (pp. 1-7). IEEE.
[3] Ranjan, J., & Malik, K. (2007). Effective educational process: a data-mining approach. Vine.
[4] Stiggins, R. J. (1994). Student-centered classroom assessment. New York: Merrill.
[5] Siahaan, A. P. U., Djanggih, H., & Aswari, A. (2018). Technique for Order Preference by Similarity to Ideal Solution as Decision Support Method for Determining Employee Performance of Sales Section.
[6] Nasution, M. D. T. P., Rossanty, Y., Achmad Daengs, G. S., Sahat, S., Rosmawati, R., Kurniasih, N., ... & Kadir, I. A. (2018). Decision support rating system with Analytical Hierarchy Process method. Int. J. Eng. Technol, 7(2.3), 105-108.
[7] Banerjee, D., & Roy, B. (2017). Application of Data Mining with R Programming to Implement ERP for Digitization of Valuation Reports on Commercial Vehicles and Passenger Cars. International Journal of Management, 8(6).
[8] Kasturi, K., et.al., (2020, April). Android Application for Orphanage Management using Data Mining Techniques, Alochana Chakra Journal,9(4),340-346.
[9] Kasturi, K.,et.al., (2019). A Hybrid Ensemble Classification Approach to Determine the Impact of Asthma in Association with Gastro Esophageal Reflux Symptoms. Indian Journal of Public Health Research & Development, 10(2), 191-196.
[10] Matsebula, F., & Mnkandla, E. (2017, September). A big data architecture for learning analytics in higher education. In 2017 IEEE AFRICON (pp. 951-956). IEEE.
[11] Barnes, M. (2013). Role reversal: Achieving uncommonly excellent results in the student-centered classroom. ASCD.