The Use of Study Apps among the High School Students – A Data Mining Approach

Roshni Kurian, N SaiDharshana, Rajalakshmi V R

***Abstract:** Learning apps are gaining huge popularity among high school students these days. It is becoming increasingly apparent that their impact and benefits benefit varies from subject to subject; broadly speaking, Learning Apps seem best suitable for subjects like Mathematics which emphasis formal and spatial learning rather than for subjects which emphasize verbal learning.*

This paper studies the use of learning apps among high school students and tries to quantify their impact in terms of academic benefit . It has been noted that with learning apps, users tend to look for immediate feedback on the learning process; this could result in improvements to the learning methods and efficiency. But there is increasing evidence that learning apps may not always provide the desired results and that their impact varies from subject to subject.

Here we employ Data Mining Techniques to analyse students' learning habits. Data Mining has provided elegant and efficient solutions to problems in different fields like education, medicine, business etc. The results of this study reveal and quantify the effectiveness of apps as a method of learning for a particular subject.

The algorithm used for this study is Naive Bayesian algorithm and tool used for calculating the result is Weka. Naive Bayesian algorithm is probability based.

***Keyword:** Data Mining, Learning apps, Naive Bayesian, Weka*

I. INTRODUCTION

The development of mobile technology has provided us with vastly enhanced opportunities to receive education. Increased spread of mobile learning Applications has greatly reduced inequities in access to technology and its use to achieve academic progress and greater personal empowerment. The rapid spread and deployment of 4G technology has increased the number of apps downloaded by users. Increase in network speed has led to more apps being available for user download. Mobile learning is potentially a great educationaid.

For any educational institution, student performance is of highest importance. Academic performance of students is usually measured by conducting various examinations, assessments and other form of measurements. However academic performance may vary from student to student as each student has different level of performance.

Revised Manuscript Received on March 27, 2019.

Roshni Kurian, Dept. of CS & IT, Amrita Vishwa Vidyapeetham, Kochi , India.

N Sai Dharshana, Dept. of CS & IT, Amrita VishwaVidyapeetham, Kochi , India.

Rajalakshmi VR, Dept. of CS & IT, Amrita Vishwa Vidyapeetham, Kochi, India.

Most of theeducational institutions are increasingly asking their students to make use of the learning apps available onlineto improve their performance. Our major aim is to identify whether the usage of these learning apps are actually beneficial to the students – after all, their parents need to bear the expenses for the purchase of these apps.Our study examines the impact of the use of learning apps by students by employing a naïve Bayesian algorithm. This is essentially, a probability theory-based approach. The naive Bayesian algorithm, which is based on the Bayes Theorem, formulates predictions by calculatingprobabilities derived from patterns in observed data. This approach is found to work well on text categorizations. Our broad conclusion from this study is that the use of learning apps is actually beneficial to the students. We also offer evidence in support of this conclusion.

II. BACKGROUNDSTUDY

The use of learning apps is beginning to have a major impact on the performance of students. We use naive Bayesian data mining technique to predict if the usage of learning apps is actually beneficial to the students and to quantify this benefit. Here, we have mainly focused on high school students to collect data in order to arrive at a conclusion.

Our focus is on identifying proper attributes and method for prediction [6].SuchitaBorkar and K. Rajeswari have taken students attendance and marks (which is further categorized) [7], Here we consider marksobtained in each subject and usage of learning apps for each subject as the majorattributes.Forprediction, we use the Naive Bayesian algorithm, which is a probability based algorithm [3].This is one of the most accurate algorithms available in data mining. It is best suitable for categorical attributes [1]. The tool used is Weka. It provides researchers with a cluster of machine learning algorithms and data processing tools[2].

III. DATA MINING PROCESS

For this study we gathered about 100 data records from a school namely, Citadel Residential School, Ranny affiliated to CBSE.The data was collected from mainly high school students i.e. students belonging to 8th and 9th standard. The data gathered was then analysed using the naïve Bayesian technique, which helped in predicting if the usage of learning apps were beneficial to the students in achieving higher scores. The following were the steps involved:

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

A. Data preparation

Initially data was collected from the students of 8th and 9th standard of Citadel Residential School, Ranny by using a questionnaire prepared for the same.

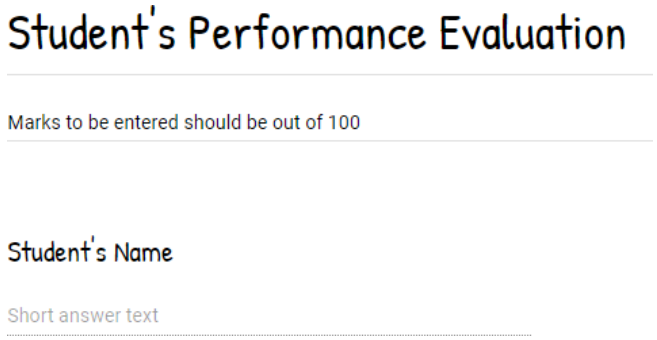


Fig. 1 Screenshot of Google form used for DataCollection

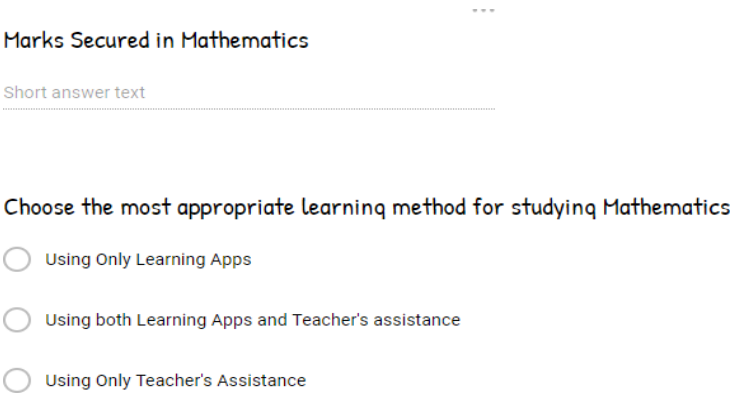


Fig. 2 Screenshot of Google Form used for DataCollecti

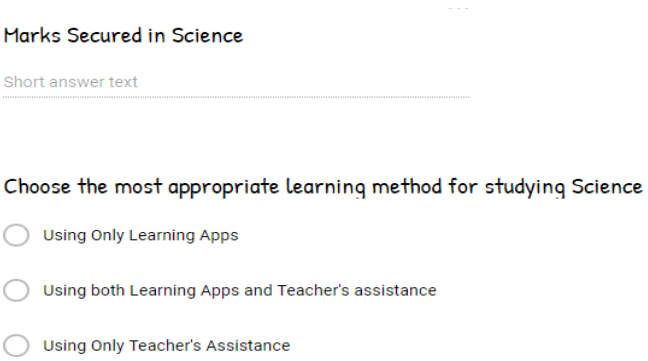


Fig. 3 Screenshot of Google Form used for DataCollection

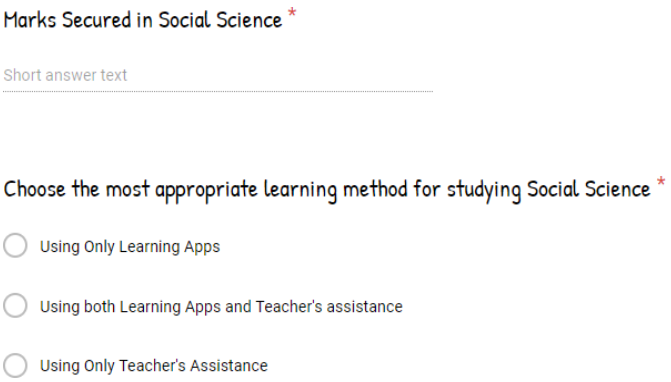


Fig. 4 Screenshot of Google Form used for DataCollection

The questionnaire was prepared in google forms and thus,we could directly get the results in .csv file format.

B. Data Selection and Transformation

Initial Dataset is created with data collected from the students. A sample of the data collected is shown in the next page.

Student's Name	Marks Secured in Mathematics	Choose the most appropriate learning method for studying	Marks Secured in Social Science	Choose the most appropriate learning method for studying Social S
DEEPA	86	Using both Learning Apps and Teacher's assistance	75	Using Only Learning Apps
Karthika .A	100	Using Only Teacher's Assistance	100	Using Only Teacher's Assistance
SANIA SUSAN SAN	90	Using Only Teacher's Assistance	82	Using Only Teacher's Assistance
ANJU ABRAHAM	52	Using Only Teacher's Assistance	48.75	Using Only Teacher's Assistance
FEBY SHOJAN	57	Using both Learning Apps and Teacher's assistance	62	Using Only Teacher's Assistance
ALIYA FATHIMA	77	Using Only Teacher's Assistance	62	Using Only Teacher's Assistance
REEMA SUSAN	48.75	Using Only Teacher's Assistance	45	Using Only Teacher's Assistance
DIYA ANN JOBY	70	Using Only Teacher's Assistance	51	Using Only Teacher's Assistance
ROHAN ABRAHAN	94	Using Only Teacher's Assistance	86	Using Only Teacher's Assistance
JOB GEORGE JAME	34	Using Only Teacher's Assistance	47.5	Using Only Learning Apps
ROHAN GEORGE	88	Using both Learning Apps and Teacher's assistance	83	Using both Learning Apps and Teacher's assistance
S JESSIE SHARON	50	Using Only Teacher's Assistance	60	Using both Learning Apps and Teacher's assistance
AFZALKHAN A S	47.5	Using Only Teacher's Assistance	40	Using Only Teacher's Assistance
ALEPH JOHN	35	Using Only Teacher's Assistance	32	Using Only Teacher's Assistance
REGAL K MATHEW	45	Using Only Teacher's Assistance	40	Using Only Teacher's Assistance
BLESSON MATHEV	70	Using Only Teacher's Assistance	61	Using Only Teacher's Assistance
ABIN SURESH TON	50	Using both Learning Apps and Teacher's assistance	60	Using Only Learning Apps
SIDDHARTH PRAS	82.5	Using Only Teacher's Assistance	53.75	Using Only Teacher's Assistance
ABISH SELVEN	68.75	Using both Learning Apps and Teacher's assistance	55	Using both Learning Apps and Teacher's assistance
AUGUSTIN SABU	95	Using Only Teacher's Assistance	86	Using both Learning Apps and Teacher's assistance
NEVEN JOSEPH	57	Using Only Teacher's Assistance	25	Using Only Teacher's Assistance
ABEL TOM MATHE	68.75	Using both Learning Apps and Teacher's assistance	98	Using Only Teacher's Assistance
JENCY ANN	57.5	Using Only Teacher's Assistance	41.87	Using both Learning Apps and Teacher's assistance

Fig. 5 Screenshot of the .csv file

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

A new training data set was created by categorizing the student marks into three different groups as follows:

Very Good	Above 80
Average	Below 80
Low	Below 40

App Usage was also categorized as fair and poor. Here, by 'fair' we mean both the app and teacher's assistance were available and 'poor' indicates a dependence only on the teacher's guidance.

The new training data set is formulated with threefields: Subject, Marks and App Usage, with the above mentioned categorization.

subject	marks	app-usage
maths	very good	fair
maths	very good	poor
maths	very good	poor
maths	avg	poor
maths	avg	fair
maths	avg	poor
maths	avg	poor
maths	avg	poor
maths	very good	poor

Fig. 6 Screenshot of new dataset

science	avg	fair
science	avg	poor
science	avg	fair
science	avg	fair
science	avg	poor
science	avg	poor
science	'very good'	fair
science	'very good'	fair

Fig. 7 Screenshot of new dataset

'social science'	'very good'	poor
'social science'	avg	poor
'social science'	avg	fair
'social science'	avg	poor
'social science'	avg	poor
'social science'	avg	poor
'social science'	'very good'	poor

Fig. 8 Screenshot of new dataset

C. MiningModel

The Mining Model implemented here is Naïve Bayesian Model. It is a stable algorithm. It is based on conditional probabilities. It is machine learning algorithm primarily used for text based categorization which involves high dimensional training data sets. It helps in making quick predictions. It uses Bayes'Theorem.

The Theorem is given by the equation:

P(A|B) =

Where P(A|B) : Conditional probability of occurrence of event A given event B is true.

P(A) and P(B) :Probabilities of the occurrence of the events A and B respectively.

P(B|A): probability of occurrence of the event B given event A istrue.

Basic Working of the algorithmConvert the dataset into frequent table

Create likelihood table by finding the probabilities of each occurrence.

Nave Bayesian equation to calculate the posterior probability for each class. The class with the highest posterior probability is the outcome ofprediction.

D. Tool

The tool used for the study is Weka - Weka stands for Waikato Environment for Knowledge Analysis.

It has several graphical user interfaces that enable easy access. Data can be uploaded from various sources and in different formats including .csv, ARFF, etc

E. Result

From this study we can easily predict the 'App Usage' for a particular subject. From the data collected we got to know that the use of learning apps is beneficial for Maths and Science but not for Social Science.

The Results and inputs have been shown below:

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

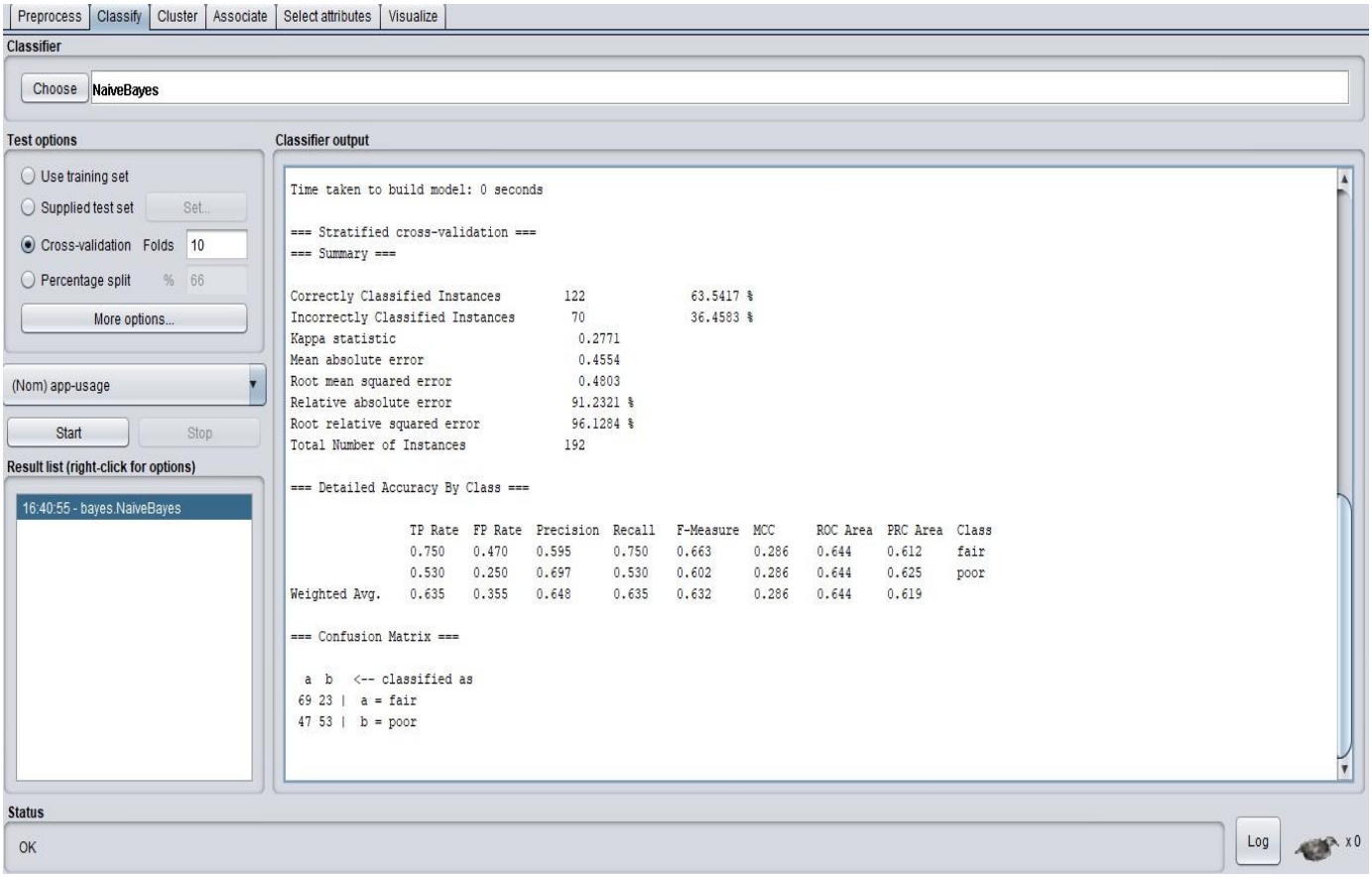
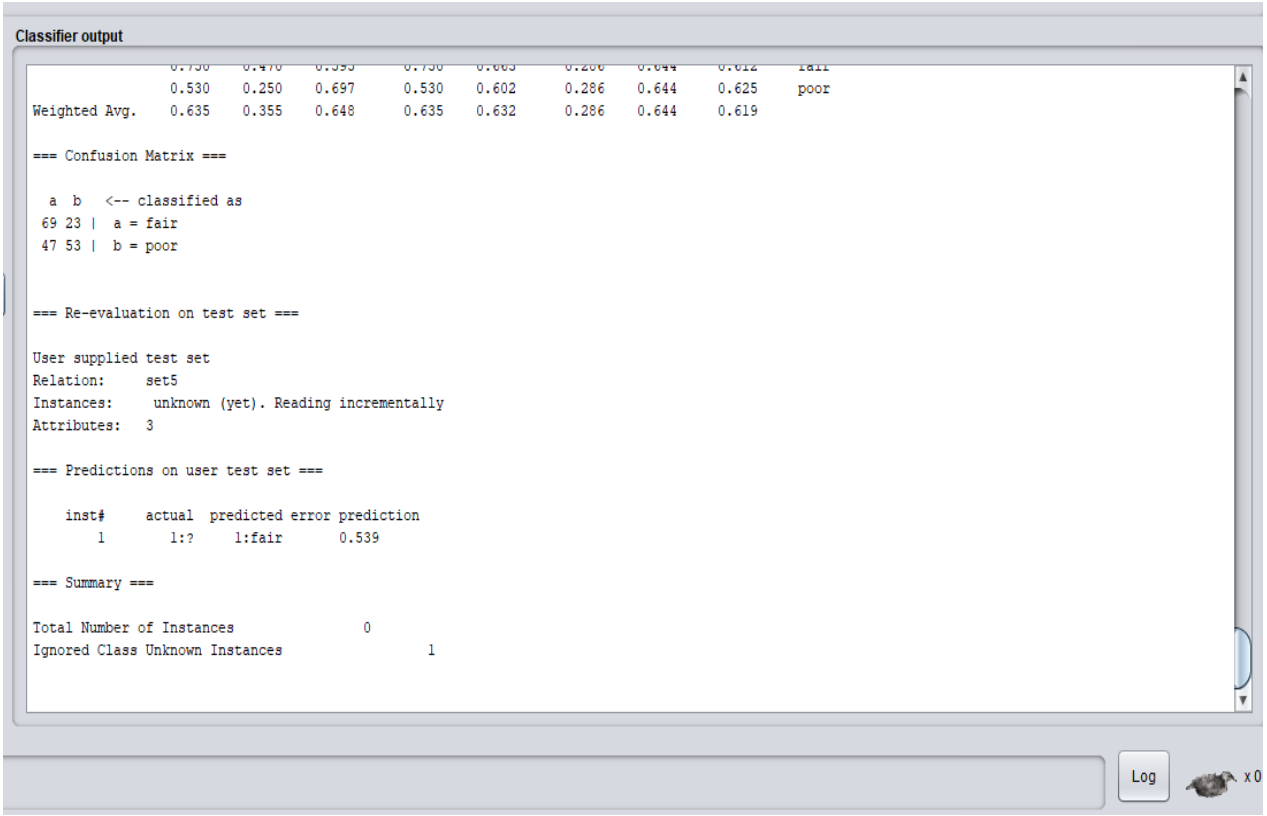


Fig 9 : Classifier Output

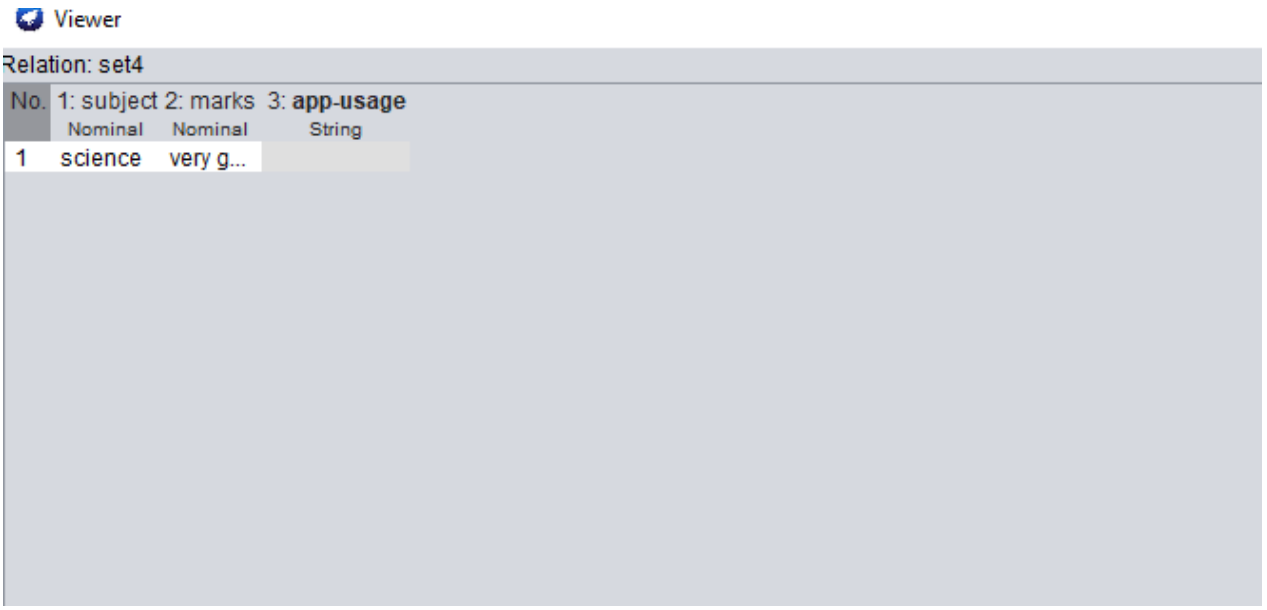
Fig 10 : Science Average Training Set



```
Classifier output

                 0.530    0.250    0.697      0.530    0.602      0.286    0.644     0.625     poor
Weighted Avg.    0.635    0.355    0.648      0.635    0.632      0.286    0.644     0.619

=== Confusion Matrix ===

  a  b   <-- classified as
 69 23 |  a = fair
 47 53 |  b = poor


=== Re-evaluation on test set ===

User supplied test set
Relation:     set5
Instances:     unknown (yet). Reading incrementally
Attributes:   3

=== Predictions on user test set ===

    inst#     actual  predicted error prediction
        1        1:?     1:fair       0.539

=== Summary ===

Total Number of Instances                0
Ignored Class Unknown Instances                  1
```

Log x 0

Fig 11 : Science –Average Output

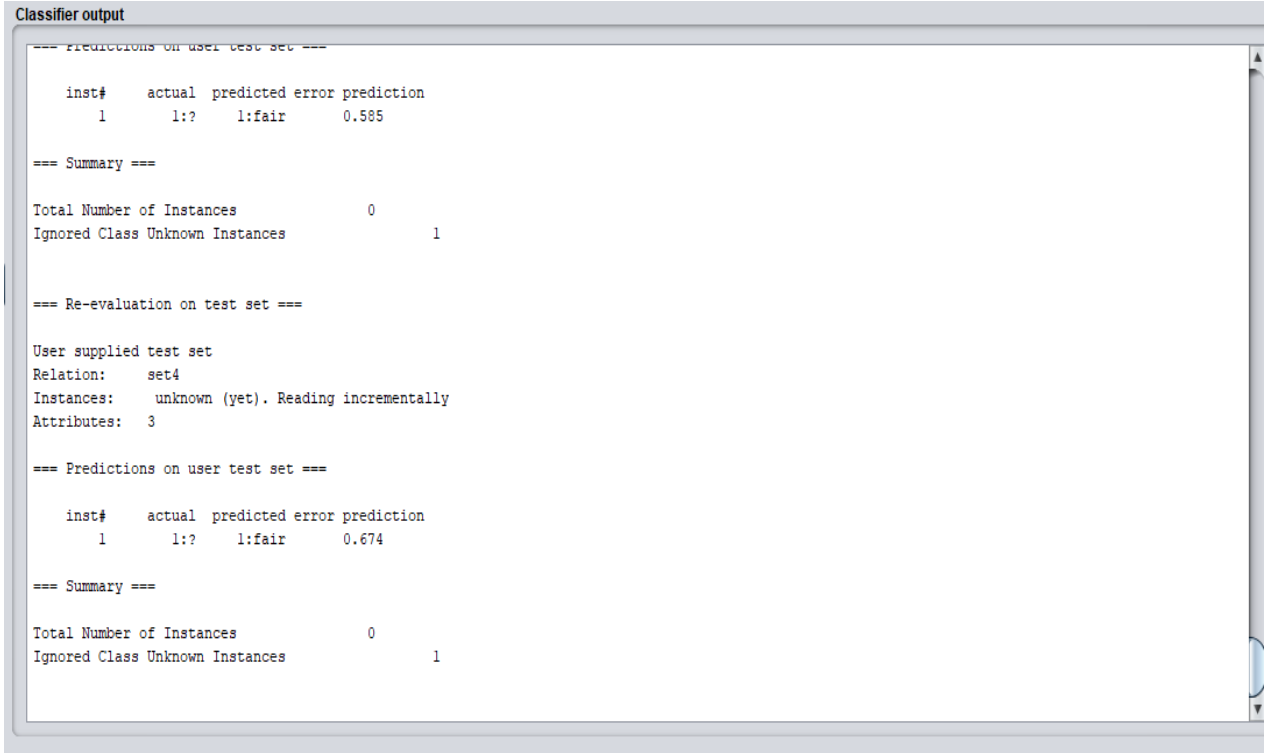


Viewer

Relation: set4

No.	1: subject	2: marks	3: app-usage
	Nominal	Nominal	String
1	science	very g...	

Fig 12 : Science – Very Good Training Set



Classifier output

```
=== Predictions on user test set ===

    inst#     actual  predicted error prediction
        1        1:?     1:fair       0.585

=== Summary ===

Total Number of Instances                0
Ignored Class Unknown Instances                  1


=== Re-evaluation on test set ===

User supplied test set
Relation:     set4
Instances:     unknown (yet). Reading incrementally
Attributes:   3

=== Predictions on user test set ===

    inst#     actual  predicted error prediction
        1        1:?     1:fair       0.674

=== Summary ===

Total Number of Instances                0
Ignored Class Unknown Instances                  1
```

Fig 13 : Science – Very GoodOutput

Viewer

Relation: set7

No.	1: subject	2: marks	3: app-usage
	Nominal	Nominal	String
1	social ...	avg	

Fig 14 : Social –Average TrainingSet

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

Classifier output

```
                 0.530    0.250    0.697      0.530    0.602      0.286    0.644     0.625     poor
Weighted Avg.    0.635    0.355    0.648      0.635    0.632      0.286    0.644     0.619

=== Confusion Matrix ===

  a  b   <-- classified as
 69 23 |  a = fair
 47 53 |  b = poor


=== Re-evaluation on test set ===

User supplied test set
Relation:     set7
Instances:     unknown (yet). Reading incrementally
Attributes:   3

=== Predictions on user test set ===

    inst#     actual  predicted error prediction
        1        1:?     2:poor       0.702

=== Summary ===

Total Number of Instances                0
Ignored Class Unknown Instances                  1
```

Fig 15 : Social – Average Output

Viewer

Relation: set6

No.	1: subject	2: marks	3: app-usage
	Nominal	Nominal	String
1	social ...	very g...	

Fig 16 : Social – Very Good Training Set

Classifier output

```
                 0.750    0.470    0.595      0.750    0.663      0.286    0.644     0.612     fair
                 0.530    0.250    0.697      0.530    0.602      0.286    0.644     0.625     poor
Weighted Avg.    0.635    0.355    0.648      0.635    0.632      0.286    0.644     0.619

=== Confusion Matrix ===

  a  b   <-- classified as
 69 23 |  a = fair
 47 53 |  b = poor


=== Re-evaluation on test set ===

User supplied test set
Relation:     set6
Instances:     unknown (yet). Reading incrementally
Attributes:   3

=== Predictions on user test set ===

    inst#     actual  predicted error prediction
       1        1:?     2:poor       0.572

=== Summary ===

Total Number of Instances                 0
Ignored Class Unknown Instances                 1
```

Fig 17 : Social – Very Good Output

Viewer

Relation: set3

No.	1: subject	2: marks	3: app-usage
	Nominal	Nominal	String
1	maths	avg	

Fig 18 :Maths – Average Training Set

Classifier output

```
                 0.530    0.250    0.697      0.530    0.602      0.286    0.644     0.625     poor
Weighted Avg.    0.635    0.355    0.648      0.635    0.632      0.286    0.644     0.619

=== Confusion Matrix ===

  a  b   <-- classified as
 69 23 |  a = fair
 47 53 |  b = poor


=== Re-evaluation on test set ===

User supplied test set
Relation:     set3
Instances:     unknown (yet). Reading incrementally
Attributes:   3

=== Predictions on user test set ===

    inst#     actual  predicted error prediction
        1        1:?     1:fair       0.585

=== Summary ===

Total Number of Instances                0
Ignored Class Unknown Instances                  1
```

Fig 19 :Maths – Average Output

Relation: set2

No.	1: subject	2: marks	3: app-usage
	Nominal	Nominal	String
1	maths	very g...	

Fig 20 :Maths – Very Good Training Set

Classifier output

```
                 0.530    0.250    0.697      0.530    0.602      0.286    0.644     0.625     poor
Weighted Avg.    0.635    0.355    0.648      0.635    0.632      0.286    0.644     0.619

=== Confusion Matrix ===

  a  b   <-- classified as
 69 23 |  a = fair
 47 53 |  b = poor


=== Re-evaluation on test set ===

User supplied test set
Relation:     set2
Instances:     unknown (yet). Reading incrementally
Attributes:   3

=== Predictions on user test set ===

    inst#     actual  predicted error prediction
       1        1:?     1:fair       0.713

=== Summary ===

Total Number of Instances                0
Ignored Class Unknown Instances                 1
```

Fig 21 : Maths – Very Good Output

The output for Mathematics and Science was predicted as 'fair' while for Social Science it was predicted as 'poor'. That is, For Social Science, majority of the students who secured good marks took only teacher's assistance for their study while for Math and Science, students securing good marks made use of the learning Apps.

IV. CONCLUSION

A clear conclusion emerges from our study: learning Apps can substantially enhance the student's assimilation of more analytical and formal subjects like Mathematics and Science – indeed, these are disciplines where interactive learning and discussions tend to give stronger benefits than solitary study; however, for more descriptive and verbal intelligence-intensive areas like Social sciences, the impact and benefits of using learning Apps tend to be much thinner. These clear findings could be further quantifiedto provide guidance while formulating academic programs for the younger generation.

REFERENCE

1 MOKHAIRI MAKHTAR, 2 HASNAH NAWANG, 3 SYADIAH NOR WAN SHAMSUDDIN," ANALYSIS ON STUDENTS PERFORMANCE USING NAÏVE BAYES CLASSIFIER", Journal of Theoretical and Applied InformationTechnology

2 Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer Peter Reutemann, Ian H. Witten, Mark Hall," The WEKA Data Mining Software: An Update", SIGKDDExplorations

3 Naive Bayes classifiers, Available:https://en.wikipedia.org/wiki/Naive_Bayes_classifier

4 Weka,https://en.wikipedia.org/wiki/Weka_(machine_learning)

5 Chapter 10 "Introduction to Weka", Data Mining: Practical Machine Learning Tools and Techniques (Third Edition) by Ian H.Witten,EibeFrank, and Mark A.Hall

6 A. Mohamed, W. Husain, and A. Rashid, "The Third Information SystemsInternationalConferenceAReviewonPredictingStudent's Performance using Data Mining Techniques," Procedia - ProcediaComput. Sci., vol. 72, pp. 414–422,2015

7 SuchitaBorkar, K. Rajeswari," Predicting Students Academic Performance Using Education Data Mining", International Journal of Computer Science and Mobile Computing,2013

8 M. M. A. Tair, A. M. El-Halees, Mining educational data to improve students' performance: a case study, International Journal of Information.

9 Ali Buldua, KeremÜçgün, Data mining application on students' data. Procedia Social andBehavioral
Sciences 2 5251–5259, 2010.

10 Singh, Randhir. An Empirical Study of Applications of Data Mining Techniques for Predicting Student
Performance in Higher Education, 2013.

11 Anwar, M. A.,and Naseer Ahmed. "Knowledge Mining in Supervised and UnsupervisedAssessment
Data of Students' Performance." 2011 2nd International Conference on Networking and Information
Technology IPCSIT vol. Vol. 17. 2011.

12 R. V. Monika Goyal, "Applications of Data Mining in Higher Education," IJCSI Int. J. Comput. Sci. Issues, vol. 9, no. 2, pp. 113–120,2012

13 A. El-Halees, "Mining Students Data To Analyze Learning Behaviour : a Case Study Educational Systems," Work, no. February, 2008.

14 Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., &Witten, I. H. (2009). The WEKA data mining software: an update. ACM SIGKDD explorations newsletter, 11(1),10-18.

15 Dai, Y., & Sun, H. (2014). The naive Bayes text classification algorithm based on rough set in the cloud platform, Journal of Chemical and Pharmaceutical Research, 2014, 6(7): 1636-1643.

AUTHORS PROFILE

Roshni Kurian, PG Student, Master of Computer Applications, Amrita VishwaVidyapeetham School of Arts and Sciences, Brahmasthanam, Edappally North P.O. Kochi – 682 024,Kerala.

N SaiDharshana, PG Student, Master of Computer Applications, Amrita VishwaVidyapeetham School of Arts and Sciences, Brahmasthanam, Edappally North P.O. Kochi – 682024,Kerala.

Rajalakshmi V R, Assistant Professor, Dept. of CS & IT, Amrita VishwaVidyapeetham, Amrita School of Arts and Sciences, Edappally North P.O. Kochi- 682024, Kerala. Quallification -M.sc (Information Technology) M.Tech (Information Technology). Research papersCOMPARATIVE STUDY OF MOBILE AD HOC NETWORK ROUTINGPROTOCOLS, An extension to page rank algorithm, A New Approach to Increase the Storage Efficiency of Databases Using BIT Representation, An Effective Approach to Improve Storage Efficiency Using Variable bit Representation. Area of interest-Block chains,Bigdata, Cloud Computing, Database.