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 vary from subject to subject; broadly speaking, Learning Apps seem best suited 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 education aid.

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.

Most of the educational institutions are increasingly asking their students to make use of the learning apps available online to 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 naïve Bayesian algorithm, which is based on the Bayes Theorem, formulates predictions by calculating probabilities 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. BACKGROUND STUDY

The use of learning apps is beginning to have a major impact on the performance of students. We use naïve 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]. Suchita Borkar and K. Rajeswari have taken students’ attendance and marks (which is further categorized) [7]. Here we consider marks obtained in each subject and usage of learning apps for each subject as the major attributes. For prediction, 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:

Revised Manuscript Received on March 27, 2019.
Roshni Kurian, Dept. of CS & IT, Amrita Vidyapeetham, Kochi, India.
N Sai Dharshana, Dept. of CS & IT, Amrita Vidyapeetham, Kochi, India.
Rajalakshmi VR, Dept. of CS & IT, Amrita Vidyapeetham, Kochi, India.
A. Data preparation
Initially data was collected from the students of 8\textsuperscript{th} and 9\textsuperscript{th} standard of Citadel Residential School, Ranny by using a questionnaire prepared for the same.

**Fig. 1 Screenshot of Google form used for DataCollection**

**Student's Performance Evaluation**

Mark to be entered should be out of 100

<table>
<thead>
<tr>
<th>Student's Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>Short answer text</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Marks Secured in Science</th>
</tr>
</thead>
<tbody>
<tr>
<td>Short answer text</td>
</tr>
</tbody>
</table>

**Choose the most appropriate learning method for studying Science**

- Using Only Learning Apps
- Using both Learning Apps and Teacher's assistance
- Using Only Teacher's Assistance

**Fig. 2 Screenshot of Google Form used for DataCollection**

<table>
<thead>
<tr>
<th>Marks Secured in Mathematics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Short answer text</td>
</tr>
</tbody>
</table>

**Choose the most appropriate learning method for studying Mathematics**

- Using Only Learning Apps
- Using both Learning Apps and Teacher's assistance
- Using Only Teacher's Assistance

**Fig. 3 Screenshot of Google Form used for DataCollection**

<table>
<thead>
<tr>
<th>Marks Secured in Social Science</th>
</tr>
</thead>
<tbody>
<tr>
<td>Short answer text</td>
</tr>
</tbody>
</table>

**Choose the most appropriate learning method for studying Social Science**

- Using Only Learning Apps
- Using both Learning Apps and Teacher's assistance
- Using Only Teacher's Assistance

**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.

<table>
<thead>
<tr>
<th>Student's Name</th>
<th>Marks Secured in Mathematics</th>
<th>Choose the most appropriate learning method for studying Mathematics</th>
<th>Choose the most appropriate learning method for studying Social Science</th>
</tr>
</thead>
<tbody>
<tr>
<td>DEEPA</td>
<td>86 Using both Learning Apps and Teacher’s assistance</td>
<td>75 Using Only Learning Apps</td>
<td>75 Using Only Learning Apps</td>
</tr>
<tr>
<td>Karthika A</td>
<td>100 Using Only Teacher's Assistance</td>
<td>100 Using Only Teacher’s Assistance</td>
<td>100 Using Only Teacher’s Assistance</td>
</tr>
<tr>
<td>SANIA SUSAN SAI</td>
<td>90 Using Only Teacher’s Assistance</td>
<td>82 Using Only Teacher’s Assistance</td>
<td>82 Using Only Teacher’s Assistance</td>
</tr>
<tr>
<td>ANNU ABRAHAM</td>
<td>52 Using Only Teacher’s Assistance</td>
<td>46.75 Using Only Teacher’s Assistance</td>
<td>46.75 Using Only Teacher’s Assistance</td>
</tr>
<tr>
<td>FEBY SHIJAN</td>
<td>57 Using both Learning Apps and Teacher’s assistance</td>
<td>62 Using Only Teacher’s Assistance</td>
<td>62 Using Only Teacher’s Assistance</td>
</tr>
<tr>
<td>ALIYA FATHIMA</td>
<td>77 Using Only Teacher’s Assistance</td>
<td>62 Using Only Teacher’s Assistance</td>
<td>62 Using Only Teacher’s Assistance</td>
</tr>
<tr>
<td>REEMA SUSAN</td>
<td>48.75 Using Only Teacher’s Assistance</td>
<td>45 Using Only Teacher’s Assistance</td>
<td>45 Using Only Teacher’s Assistance</td>
</tr>
<tr>
<td>DIYA ANN JIBY</td>
<td>70 Using Only Teacher’s Assistance</td>
<td>51 Using Only Teacher’s Assistance</td>
<td>51 Using Only Teacher’s Assistance</td>
</tr>
<tr>
<td>ROHAN ABRAHAM</td>
<td>94 Using Only Teacher’s Assistance</td>
<td>86 Using Only Teacher’s Assistance</td>
<td>86 Using Only Teacher’s Assistance</td>
</tr>
<tr>
<td>JOB GEORGE JANI</td>
<td>34 Using Only Teacher’s Assistance</td>
<td>47.5 Using Only Learning Apps</td>
<td>47.5 Using Only Learning Apps</td>
</tr>
<tr>
<td>ROHAN GEORGE</td>
<td>88 Using both Learning Apps and Teacher’s assistance</td>
<td>83 Using both Learning Apps and Teacher’s assistance</td>
<td>83 Using both Learning Apps and Teacher’s assistance</td>
</tr>
<tr>
<td>S JESSIE SHARON</td>
<td>50 Using Only Teacher’s Assistance</td>
<td>60 Using both Learning Apps and Teacher’s assistance</td>
<td>60 Using both Learning Apps and Teacher’s assistance</td>
</tr>
<tr>
<td>AFZAL KHAN A S</td>
<td>47.5 Using Only Teacher’s Assistance</td>
<td>40 Using Only Teacher’s Assistance</td>
<td>40 Using Only Teacher’s Assistance</td>
</tr>
<tr>
<td>ALEPH JOHN</td>
<td>35 Using Only Teacher’s Assistance</td>
<td>32 Using Only Teacher’s Assistance</td>
<td>32 Using Only Teacher’s Assistance</td>
</tr>
<tr>
<td>REGAL K MATHEN</td>
<td>45 Using Only Teacher’s Assistance</td>
<td>40 Using Only Teacher’s Assistance</td>
<td>40 Using Only Teacher’s Assistance</td>
</tr>
<tr>
<td>BLESSON MATHEN</td>
<td>70 Using Only Teacher’s Assistance</td>
<td>61 Using Only Teacher’s Assistance</td>
<td>61 Using Only Teacher’s Assistance</td>
</tr>
<tr>
<td>ABIN SURESH TOI</td>
<td>50 Using both Learning Apps and Teacher’s assistance</td>
<td>60 Using Only Learning Apps</td>
<td>60 Using Only Learning Apps</td>
</tr>
<tr>
<td>SIDDHARTH PRAS</td>
<td>82.5 Using Only Teacher’s Assistance</td>
<td>58.75 Using Only Teacher’s Assistance</td>
<td>58.75 Using Only Teacher’s Assistance</td>
</tr>
<tr>
<td>ABISH SEVEN</td>
<td>68.75 Using both Learning Apps and Teacher’s assistance</td>
<td>55 Using both Learning Apps and Teacher’s assistance</td>
<td>55 Using both Learning Apps and Teacher’s assistance</td>
</tr>
<tr>
<td>AUGUSTIN SABU</td>
<td>95 Using Only Teacher’s Assistance</td>
<td>86 Using both Learning Apps and Teacher’s assistance</td>
<td>86 Using both Learning Apps and Teacher’s assistance</td>
</tr>
<tr>
<td>NEVIN JOSEPH</td>
<td>57 Using Only Teacher’s Assistance</td>
<td>25 Using Only Teacher’s Assistance</td>
<td>25 Using Only Teacher’s Assistance</td>
</tr>
<tr>
<td>ABE TCM MATEH</td>
<td>68.75 Using both Learning Apps and Teacher’s assistance</td>
<td>98 Using Only Teacher’s Assistance</td>
<td>98 Using Only Teacher’s Assistance</td>
</tr>
<tr>
<td>JENCY ANN</td>
<td>57.5 Using Only Teacher’s Assistance</td>
<td>41.67 Using both Learning Apps and Teacher’s assistance</td>
<td>41.67 Using both Learning Apps and Teacher’s assistance</td>
</tr>
</tbody>
</table>

Fig. 5 Screenshot of the .csv file
A new training data set was created by categorizing the student marks into three different groups as follows:

<table>
<thead>
<tr>
<th>Category</th>
<th>Marks</th>
</tr>
</thead>
<tbody>
<tr>
<td>Very Good</td>
<td>Above 80</td>
</tr>
<tr>
<td>Average</td>
<td>Below 80</td>
</tr>
<tr>
<td>Low</td>
<td>Below 40</td>
</tr>
</tbody>
</table>

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 three fields: Subject, Marks and App Usage, with the above mentioned categorization.

### C. Mining Model

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: $P(A|B) = \frac{P(A) P(B|A)}{P(B)}$

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 is true.

Basic Working of the algorithm:
- Convert the dataset into frequent table
- Create likelihood table by finding the probabilities of each occurrence.
- Naïve Bayesian equation to calculate the posterior probability for each class. The class with the highest posterior probability is the outcome of prediction.

### 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:
Fig 9: Classifier Output

Fig 10: Science Average Training Set
The Use of Study Apps among the High School Students – A Data Mining Approach

![Classifier output]

--- Confusion Matrix ---

a b <- classified as
69 23 | a = fair
47 30 | b = poor

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

User supplied test set
Relations: math
Instances: unknown (yet), Reading incrementally
Attributes: 3

--- Predictions on user test set ---

Last 4 actual predicted error prediction
   1 37 1:fair 0.338

--- Summary ---

Total Number of Instances 5
Ignored Class Unknown Instances 1

Fig 11: Science – Average Output

![Relation set4]

No 1: subject 2: marks 3: app_usage
Nominal Nominal String
1 science very g...

Fig 12: Science – Very Good Training Set
Fig 13 : Science – Very Good Output

Fig 14 : Social – Average Training Set
The Use of Study Apps among the High School Students – A Data Mining Approach

Fig 15 : Social – Average Output

Fig 16 : Social – Very Good Training Set
Fig 17: Social – Very Good Output

Fig 18: Maths – Average Training Set
The Use of Study Apps among the High School Students – A Data Mining Approach

--- Confusion Matrix ---

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

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

User supplied test set
Relation: set1
Instances: unknown (yet). Reading incrementally
Attributes: 9

--- Predictions on user test set ---

```
inst#    actual  predicted error prediction
  1  1.7     1.fair    0.099
```

--- Summary ---

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

Fig 19: Maths – Average Output

Fig 20: Maths – Very Good Training Set
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 quantified to provide guidance while formulating academic programs for the younger generation.

REFERENCE

2. Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer Peter Reutemann, Ian H. Witten, Mark Hall, "The WEKA Data Mining Software: An Update", SIGKDD explorations
8. M. M. A. Tair, A. M. El-Halees, Mining educational data to improve students’ performance: a case study, International Journal of Information