ANALYSIS AND PREDICTIONS ON BLENDED LEARNING READINESS AMONG INDIAN STUDENTS AT UNIVERSITIE SUSING DECISION TREE CLASSIFIER IN SCIKIT-LEARN ENVIRONMENT

Rajalakshmi R
MSc Information Technology
M.O.P Vaishnav college for women
Chennai- 600034, India
e-mail: mail2raji.cs@gmail.com

ABSTRACT

Decision trees classifiers are modest and hasty data classifiers, usually used in data mining to study the data and generate the tree and its rules that will be used to formulate predictions. One of the key challenges for knowledge discovery and data mining systems stands in developing their data analysis capability to discover out of the conventional models in data. Since the Union budget 2017, the debate around the quality of Higher Education in India has been acquiring momentum, which laid emphasis on skills development, employability and digitisation of the education process. The key lies in blended learning, a model that is fast gaining pace in the Indian context,where online tools are combined with classroom and instruction to provide an overall improvement in educational outcomes activities.The explanatory variables are students' attitude towards learning flexibility, online learning, classroom learning, degree and stream. This paper represents an implementation of the decision tree classifier algorithm using scikit-learn library for python on data collected from the survey with the purpose of predicting whether a particular student is ready for blended learning throughdecision trees.

Keywords: Data mining; Classification; Decision trees; Blended Learning; scikit-learn

I. INTRODUCTION

Data mining is a process used to extract usable data from a larger set of any raw data.Two main objectives can be distinguished in the data mining process integrated in the management system: a description objective consisting in establishing the eloquent variables and its influences; and a prediction objective.

Online educational tools are used in a wide range of context for many different goals and motives, but there is anincreasing focus on blended learning where online tools are combined with classroom and instruction to provide an overall improvement in educational outcomes activities.Students in the developing world are frequently quoted as being among the most important heirs of online education initiatives such as MOOCs.Blended learning is a hybrid form of teaching and learning which involves both classroom and online learning. Blended learning incorporates information via online courses, developed by experts from different fields, and helping students access globally developed and industry relevant course material. The approach mixes concept building and enquiry-based learning which retains human interaction in education and allows students to combine traditional classroom methods with online digital medium.

The main objective of this paper is an attempt to use data mining methodologies to study students' readiness towards blended learning. Data mining provides many tasks that could be used to study this analysis. In this research, the classification task is used to evaluate student's readiness and as there are many approaches that are used for data classification, the decision tree method is used here. An online questionnaire is used to collect data for this research and the respondents were students from different colleges in India belonging to different streams of study. The decision trees are generated using Decision tree classifier algorithm built using scikit-learn library for python in Anaconda Navigator (Spyder) Environment.

II. REVIEW OF LITERATURE

The paper [1] has presented findings of a study of mobile learning readiness among Malaysian students from two public universities. This describes about the basic readiness, skill readiness, psychological readiness, budget readiness of students from different universities. The research concludes that it is essential for educators to integrating mobile learning into academic programs and that the respondents welcomed this idea.[2] Reports on blended learning environment approach to enhance the performance of the students learning science. Students' learning experience outside the school and learning outcomes can be improved by adequately preparing for both pre- and post-visit of blended learning courses. The concept about the particular subject is strongly covered by multi-faced roles played by students and teachers. [3] Provides an insight into how the students really work and learn using technologies. Qualitative and quantitative techniques were used to gain an appreciation of the students' experience with ICT as a supporting mechanism and the blended delivery medium for the module.E-learning is becoming popular in educational institutions because ICT (information and communication technologies) [4] .Linear Regression analysis is used as a means to find that the relationship between E-learning and student is high. To find out the student's preferences, a commonly used approach is to implement a decision model that matches some relevant characteristics of the learning resources with the student's learning style. Adaptive machine learning algorithms are used as a tool to learn about the student's preferences towards E learning over time. At first, all the information available about a particular student is used to build an initial decision model based on learning styles [5].

III. SYSTEM MODEL

To build a reliable classification model, the following methodology is adopted. The methodology consists mainly of five steps: Collecting the relevant features of the problem under study, preparing the data, building the classification model, evaluating the model using one of the evaluation methods, and finally using the model for future prediction of the student performance. These steps are presented in the next subsections.

Conference Paper: International Conference on "Recent Advances in Computing and Communication"
Organized by: Department of Computer Science, SSS Shasun Jain College for Women, Chennai, India

ICT ACADEMY
Innovate... Collaborate... Educate...

33

A. Collecting the relevant features

In this step, the relevant features are collected through a survey via Google Forms and the respondents were 100 students from different colleges in India. Initially 15 attributes have been collected of which only 7 conditional attributes and one class attribute have been considered. The attributes along with their descriptions and possible values are presented in Table I. The class attribute is the student's readiness for blended learning and named BLREADY.

B. Feature selection and construction

New features were constructed from the existing features. They were Attitude towards Learning Flexibility, Attitude towards Online Learning, and Attitude towards Classroom Learning. The features were also discretized. All the respondents agreed that web

is a useful platform for learning and that technology wasn't a hindrance to anyone in any way.

C. Building the classification model

The Classification model is built using the decision tree method. The decision tree is a very good and practical model since it is comparatively fast, and can be easily converted to simple classification rules. The decision tree method relies mainly on using the information gain metric which determines the feature that is most useful. The information gain depends on the entropy measure. The categorical values of the conditional variables are encoded using one-hot encoder. The data in csv format is split into training data (70%) and test data (30%). The tree module is used to build a Decision Tree Classifier. Accuracy metrics from the predicted class variable is computed using Accuracy_score module.

Table I. Symbolic Attribute Description

Sr.No	ATTRIBUTE	DESCRIPTION	POSSIBLE VALUES	ONE-HOT ENCODED VALUES
1	Degree	Degree pursued by the student	Undergraduate, Postgraduate, Others	2,1,0
2	Stream	Stream of Study	Science, Arts, Engineering, Commerce, Others	5,0,2,1,4
3	Year	Year of Study	I, II, III, IV, V, Graduated	1,2,3,4,5,0
4	OL	Have you done any online course?	Yes, No	1,0
5	ATLF	Attitude towards Learning Flexibility	Poor, Fair, Excellent	0,1,2
6	ATOL	Attitude towards Online Learning	Poor, Fair, Excellent	0,1,2
7	ATCL	Attitude towards Classroom Learning	Poor, Fair, Excellent	0,1,2
8	BLREADY	Readiness for Blended Learning	Yes, No	1,0

1) Data Slicing

Data slicing is a step that is used to split data into train and test set. Training data set can be used specifically for our model building. Test dataset should not be mixed up while building model. We should not standardise our test set even during standardisation.

2) Decision Tree Training

DecisionTreeClassifier(): This is the classifier function for DecisionTree defined by scikit-learn. It is the main function for implementing the algorithms. Some important parameters are:

- criterion: It defines the function to measure the quality of a split. Information gain is used as the criterion.
- splitter: It defines the strategy to choose the split at each node. Best split has been chosen.
- max_depth: The max_depth parameter denotes maximum depth of the tree.
- min_samples_leaf: The minimum number of samples required to be at a leaf

The following rules are inferred from the decision tree in Figure 1.

- If the student's attitude towards classroom learning is poor, attitude towards learning flexibility is fair or excellent and the student has done online course then the student is ready for blended learning.
- If the student's attitude towards classroom learning is fair or excellent and the student belongs to commerce or

Humanities then the student is not ready for blended learning.

- If the student's attitude towards classroom learning is poor, attitude towards learning flexibility is fair or excellent, the student has not done online course and the degree is postgraduate or undergraduate then the student is ready for blended learning.
- If the student's attitude towards classroom learning is poor, attitude towards learning flexibility is poor and the attitude towards online learning is fair or excellent then the student is ready for blended learning.
- If the student's attitude towards classroom learning is poor, attitude towards learning flexibility is poor, the attitude towards online learning is poor and the degree is postgraduate or undergraduate then the student is not ready for blended learning.

IV. PERFORMANCE ANALYSIS AND RESULTS

In order to measure the performance of a classification model on the test set, the classification accuracy or error rate are usually used for this purpose. Accuracy is the ratio of the correctly predicted data points to all the predicted data points. Accuracy as a metric helps to understand the effectiveness of our algorithm. The classification accuracy is computed from the test set where it can also be used to compare the relative performance of different classifiers on the same domain. However, in order to do so, the class labels of the test records must be known. Moreover an evaluation methodology is needed to evaluate the classification model and compute the classification accuracy. The decision tree classifier resulted in an accuracy of 78% and Naïve Bayes

Conference Paper: International Conference on "Recent Advances in Computing and Communication"
Organized by: Department of Computer Science, SSS Shasun Jain College for Women, Chennai, India

ICT ACADEMY
Innovate... Collaborate... Educate...

34

algorithm resulted in an accuracy of 72%. The sample training data is shown in Table II. The sample test data is shown in Table III.

V. BIVARIATE ANALYSIS

Bivariate Analysis finds out the relationship between two features. The association between stream and readiness for blended learning is shown in Figure 2. Figure 3 shows the scatter plot between stream and attitude towards online learning. Figure 4 shows the scatter matrix of all the features
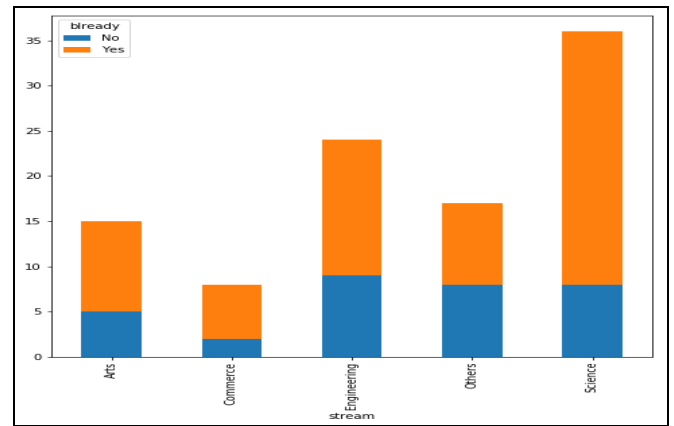
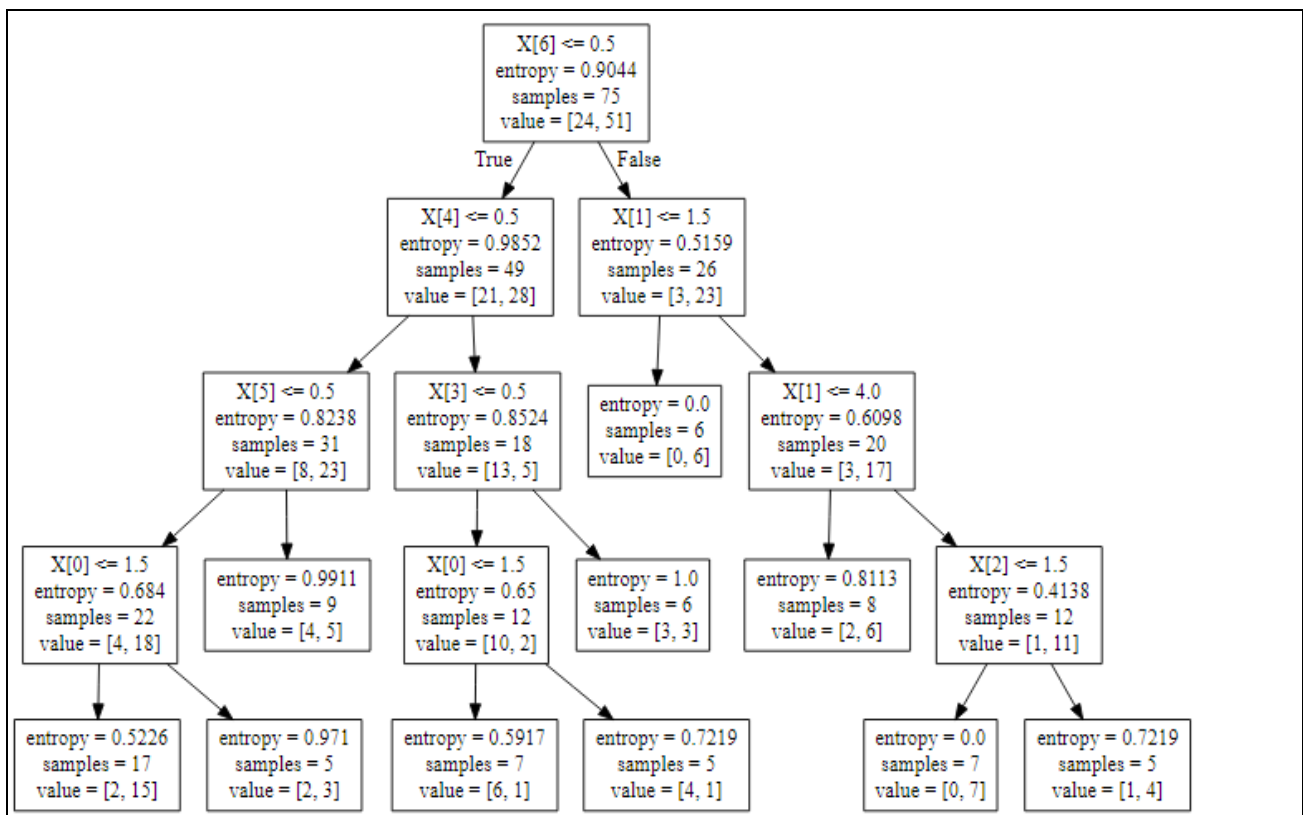Figure 2. Association between stream and blended learning readiness

Figure 1. Decision tree

.Table II. Sample Training Data

Sr.No	DEGREE	STREAM	YEAR	OL	ATLF	ATOL	ATCL	BLR
1	Postgraduate	Science	I	No	Poor	Fair	Poor	Yes
2	Postgraduate	Science	I	No	Excellent	Excellent	Excellent	Yes
3	Undergraduate	Engineering	Graduated	No	Excellent	Poor	Excellent	No
4	Postgraduate	Science	I	Yes	Excellent	Excellent	Fair	Yes
5	Undergraduate	Engineering	IV	Yes	Fair	Fair	Poor	Yes
6	Postgraduate	Science	I	No	Excellent	Excellent	Excellent	No
7	Undergraduate	Others	III	No	Excellent	Fair	Excellent	No
8	Undergraduate	Engineering	IV	No	Fair	Excellent	Excellent	No
9	Postgraduate	Science	I	Yes	Poor	Poor	Poor	Yes
10	Undergraduate	Engineering	II	No	Excellent	Excellent	Excellent	No

Conference Paper: International Conference on "Recent Advances in Computing and Communication"
Organized by: Department of Computer Science, SSS Shasun Jain College for Women, Chennai, India

ICT ACADEMY
Innovate... Collaborate... Educate...

35

Table III. Sample Test Data

| Sr. No	DEGREE	STREAM	YEAR	OL	ATLF	ATOL	ATCL	BLR
1	Undergraduate	Engineering	IV	No	Excellent	Excellent	Fair	No
2	Postgraduate	Science	II	No	Excellent	Excellent	Excellent	Yes
3	Undergraduate	Engineering	IV	No	Fair	Excellent	Excellent	Yes
4	Undergraduate	Engineering	IV	No	Excellent	Excellent	Fair	Yes
5	Postgraduate	MBA	I	Yes	Fair	Excellent	Excellent	Yes
6	Postgraduate	Engineering	IV	Yes	Excellent	Excellent	Excellent	No
7	Undergraduate	Commerce	Graduated	No	Excellent	Excellent	Poor	Yes
8	Undergraduate	Commerce	Graduated	No	Excellent	Excellent	Excellent	No
9	Postgraduate	MBA	I	No	Excellent	Excellent	Excellent	Yes
10	Undergraduate	Engineering	IV	No	Excellent	Excellent	Excellent	No
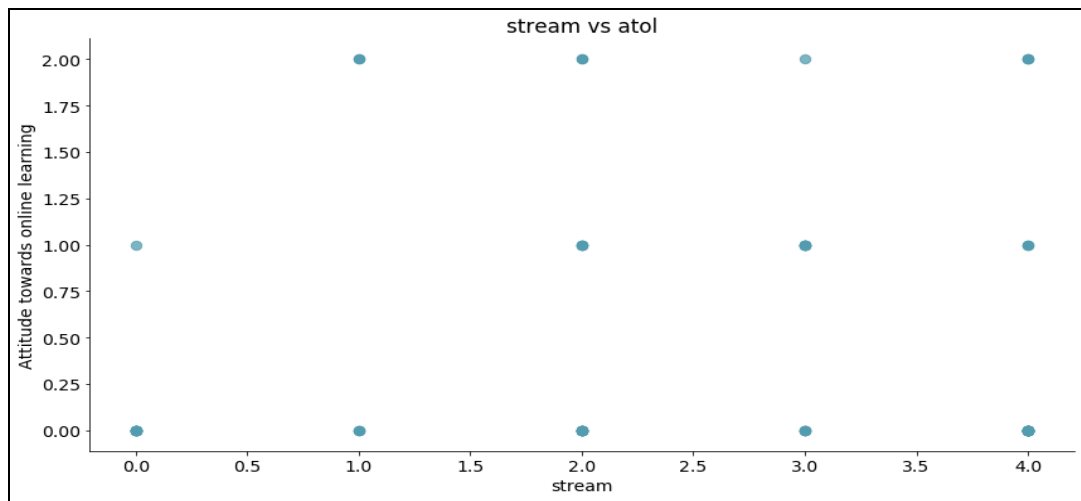
Figure 3. Scatter plot of stream Vs. attitude towards online learning
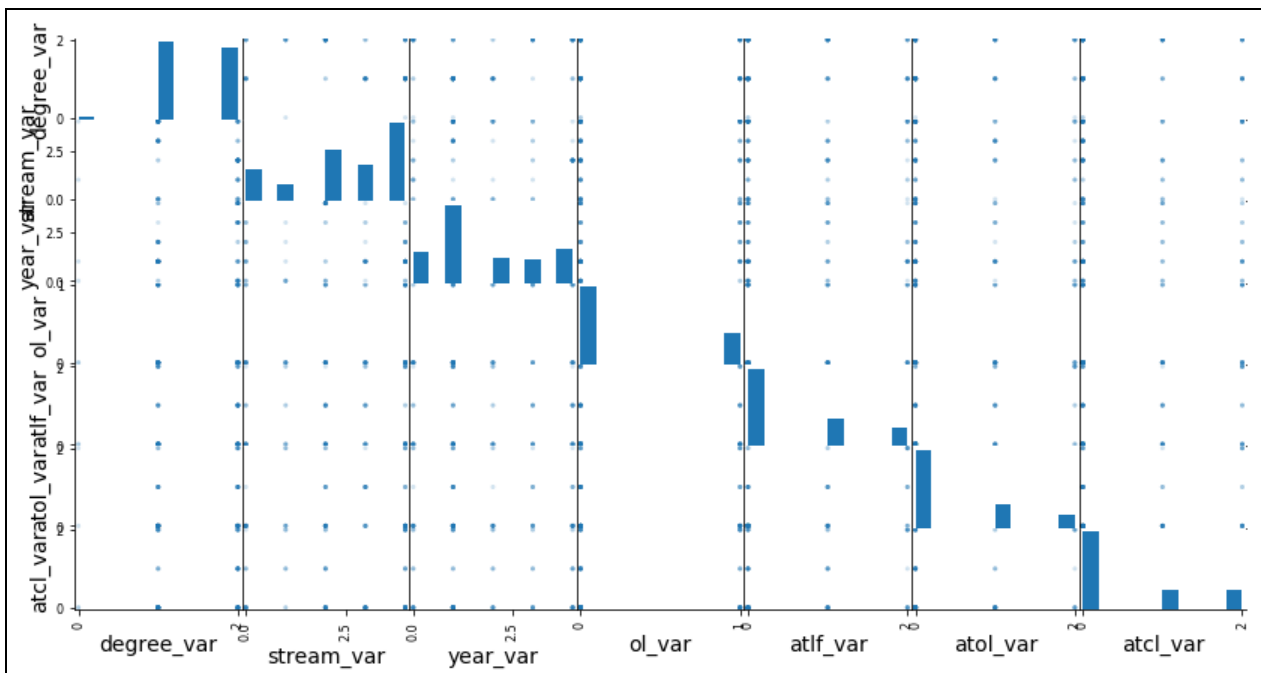

Figure 4. Scatter matrix

VI. CONCLUSION

This research is a preparatory attempt to use data mining functions to analyse and evaluate students' readiness towards blended learning. The higher managements can use such classification model to enhance the methodology of learning according to the extracted knowledge. Such knowledge can be used by the management system to improve their policies, enhance their strategies, and improve the quality of education system. One of the most striking future works is to collect a real and large data set from the students across India and apply the model using such data. Also, it could be narrowed down by taking a particular institution in hand and collecting responses from the students and thereby predicting the preference towards blended learning. Moreover, several other classification methods can also be applied to test the most suitable method that suit the structure of the data and give a better classification accuracy.

VII. REFERENCES

[1] Supyan Hussin1, Mohd Radzi Manap1, Zaini Amir1, Pramela Krish & Pramela Krish1. Mobile Learning Readiness among Malaysian Students at Higher Learning Institutes - Published by Canadian Center of Science and Education,2012

[2] Sandhya Devi Coll, Research Scholar, Curtin University, Perth, WA, Australia David Treagust. Blended Learning Environment: An Approach to Enhance Student's Learning Experiences Outside School (LEOS) –MIER Journal of Educational Studies, November 2017.

[3] Fiona Concannon, Antoinette Flynn and Mark Campbell, What campus-based students think about the quality and benefits of e-learning, British Educational Communications and Technology Agency, 2005

[4] The Impact Of E-Learning On Students Performance In Insitution , Oye, N. D.; 2A.Iahad, N., Madar, M. J. and Ab.Rahim, N., Department of Information System Universiti Teknologi Malaysia

[5] Cristina Carmona1, Gladys Castillo2, Eva Millán1, Discovering Student Preferences in E-Learning

[6] Qasem A. Al-Radaideh, Emad M. Al-Shawakfa, Mustafa I. Al-Najjar, Mining Student Data Using Decision Trees, International Arab Conference on Information Technology (ACIT'2006), Nov. 2006, Jordan.

[7] Vasile Paul Bre_felean, Babe_-Bolyai University, Faculty of Economics and Business Administration, Analysis and Predictions on Students' Behavior Using Decision Trees in Weka Environment, https://www.researchgate.net/publication/4266105 (Article in a conference proceedings)

[8] Subitha Sivakumar, Rajalakshmi Selvaraj, Predictive Modeling of Students Performance Through the Enhanced Decision Tree Advances in Electronics, Communication and Computing pp 21-36, 2017

[9] Sachin, R.B., Vijay, M.S.: A survey and future vision of data mining in educational field. In: 2012 Second International Conference on Advanced Computer Engineering & Communication Technologies, pp. 96–100 (2012)

[10] Norwawi, N.M., Abdusalam, S.F., Hibadullah, C.F., Shuaibu, B.M.: Classification of students performance in computer programming course according to learning style. In: 2nd Conference on Data Mining and Optimization, pp. 37–41 (2009)