Research paper

TO COMPARE THE DECISION TREE AND RANDOM FOREST ALGORITHM IN CATEGORIZING LEARNERS OF THE NAVI MUMBAI REGION

Ms. Nutan Sawant

Assistant Professor, Department of Information Technology, SIES (Nerul) College of Arts, Science and Commerce Navi Mumbai, India

parabnutan@gmail.com

ABSTRACT

With an increasing number of learners in the institute, it will be very difficult to correctly divide them into different categories based on their capabilities and provide them with perfect pedagogy as per their needs. To provide a smart teaching-learning process, we can categorize the learners at the beginning of the academic year based on their performance in preliminary tests. It can be categorized based on the learner's grasping skills as slow, less advanced, and advanced learners.

We can use artificial intelligence techniques to assist the institute in the classification of learners in an intelligent manner. Different Classification Algorithms of Machine Learning can be used to split the learners into different predefined classes. Based on learners' performance in the preliminary assessments, we can provide them with additional online courses along with their regular curriculum to improve their learning.

Keywords: Initial level test, Machine Learning, Classification algorithms

INTRODUCTION

Different learners come together to learn in a class. In offline classrooms, it's not easy to apply a variety of pedagogies in accordance with the needs of the learners. Instead, by using the smart Teaching-Learning approach, we can provide them with a better opportunity to improve their learning. Teachers can gather learners' preliminary assessment data, and we can cleverly classify them into various predefined groups based on certain standards using AI techniques. We can group them into slow learners, less advance learners, and advanced learners.

Machine Learning

By using Artificial Intelligence, applications can do the task on their own without human intervention. Machine learning is a part of Artificial Intelligence where we can analyze, visualize, classify and forecast data.

In Machine Learning there are supervised and unsupervised machine learning algorithms. In the case of supervised learning initially, we need to provide a training data set that consists of many features, and then based on the previous knowledge machine can predict the output. In the case of Unsupervised Learning, we can use Clustering and dimensionality reduction mechanisms to predict the output.

OBJECTIVE

The objective of this paper is to compare the classification algorithms of machine learning to smartly classify learners into different categories to enhance their learning process.

METHODOLOGY

For this paper, the questionnaire has been used as a data collection procedure. The questionnaire consists of different attributes related to their stream and logical reasoning.

947 data entries were collected from various programs.

Table No. 1: Department-wise data

Department	Number of Students
BAMMC	116
Bcom	385

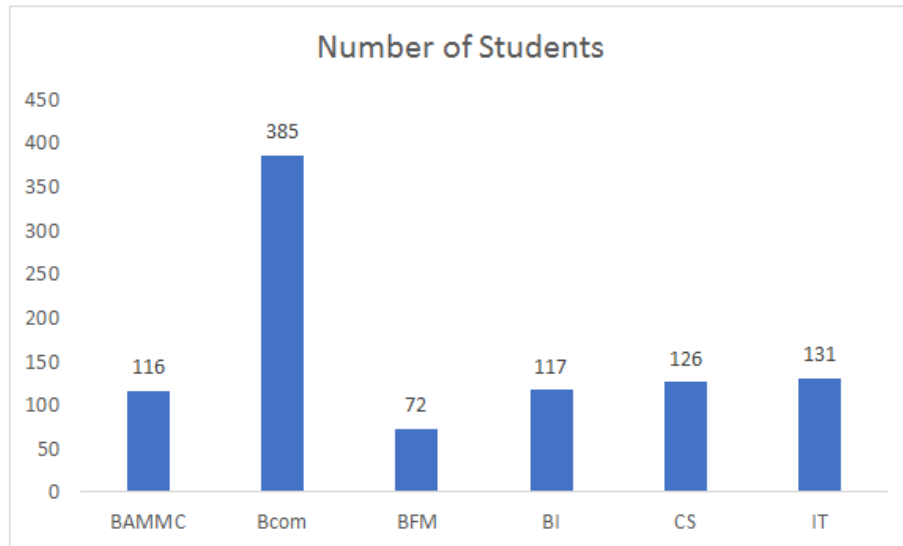|	
BFM	72
BI	117
CS	126
IT	131
Grand Total	947

Fig No. 1 Number of students

HYPOTHESIS

H0: The Random Forest algorithm works well as compared to the Decision Tree algorithm for this dataset.

H1: The Random Forest algorithm does not work well as compared to the Decision Tree algorithm for this data set.

Paper Findings

The following different classes are created based on certain criteria where entry-level test and HSC marks are considered. Based on gender below data received.

Table No. 2: Gender wise classification

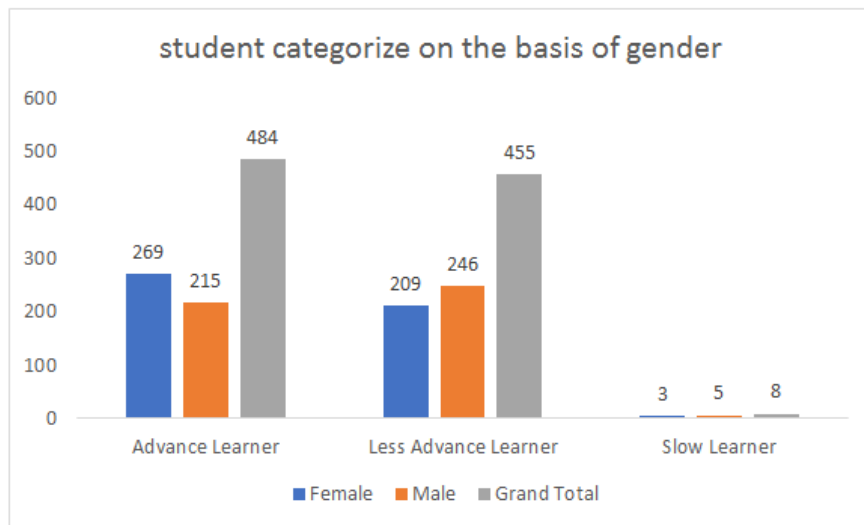| Gender	Advance Learner	Less Advance Learner	Slow Learner
Female	269	209	3
Male	215	246	5
Grand Total	484	455	8

Fig No.2 Gender wise classification

The following observations have been noted in different programs.

Table No. 3 Department-wise classification

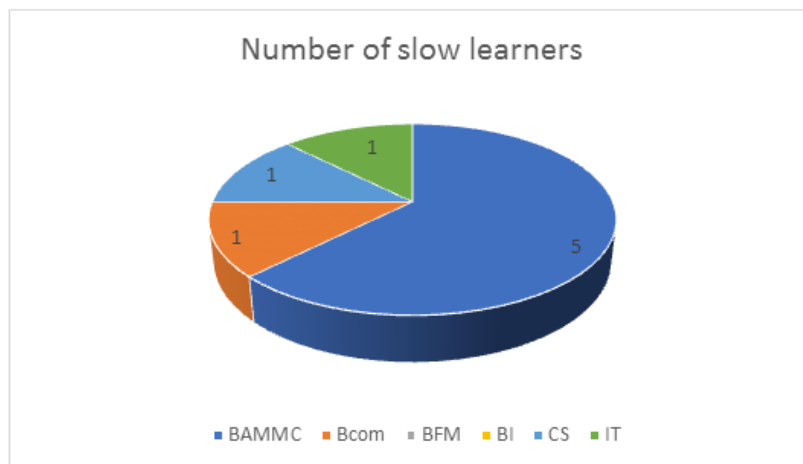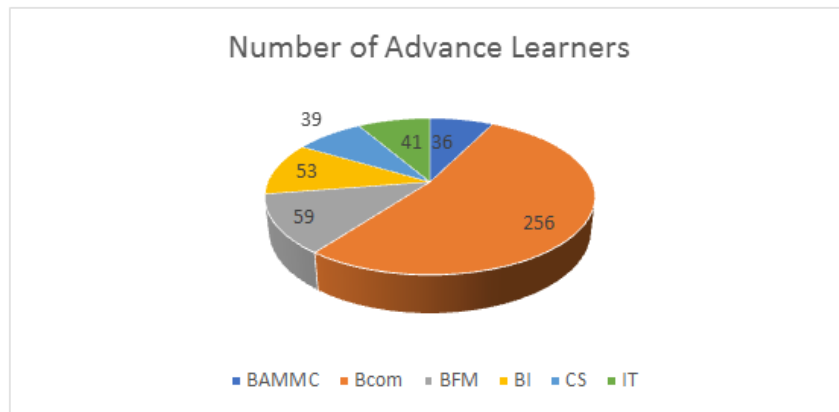| Department	Advance Learner	Less Advance Learner	Slow Learner
BAMMC	36	75	5
Bcom	256	128	1
BFM	59	13	
BI	53	64	
CS	39	86	1
IT	41	89	1
Grand Total	484	455	8

Fig No. 3 Number of slow learners

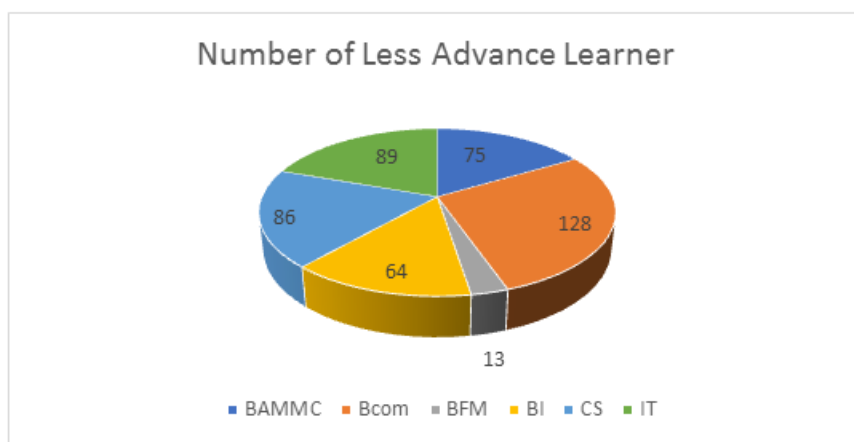**Fig No. 4** Number of advanced learners

Fig No. 5 Number of Less advanced learners

Classification of Learners using Machine Learning Algorithm

947 records were collected as a dataset.75% of the data is decided as training data set and 25% of the data is decided to be used for testing purposes. Initially, in the training phase of system development, the system is trained by using training data set which included Learners' details along with their academic progress data. Learners' data is provided to the system which acts as independent variables. Based on their preliminary test, learners are labeled under different categories. A category is a dependent variable. Once the system is trained it went to the testing phase. The tested machine can be used in the future to classify learners into different categories without human intervention. [2]

The following pair plot was received as an output of the dataset which helped in selecting the classification algorithm.
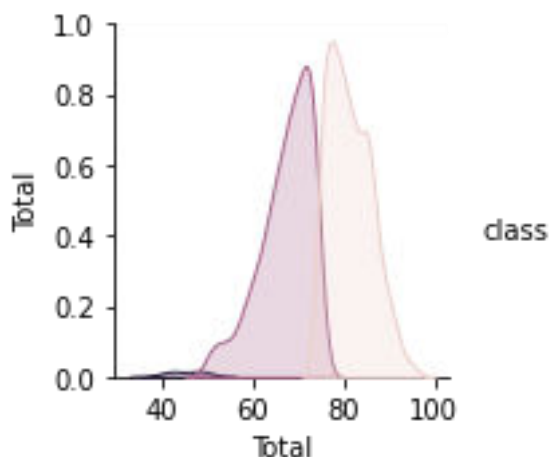
Fig No. 6 : Pairplot of dataset

The decision Tree algorithm is the simplest and easy to understand algorithm which mimics a human brain. A decision tree combines some decisions, whereas a random forest combines several decision trees. Thus, it is a long process, yet slow.

Whereas, a decision tree is fast and operates easily on large data sets, especially linear ones. The random forest model needs rigorous training.[3]

The System is tested on both algorithms and finds the accuracy of the data set.

Following confusion matrix received after applying decision tree algorithm.

[[0 128   0]

[0 108   0]

[0   0   1]]

Following confusion matrix received after applying random forest algorithm.

[[128   0   0]

[0 108   0]

[0   0   1]]

RESULT

The following Accuracy score is obtained for the dataset.

Algorithm	Accuracy score
decision tree	0.459915611814346
random forest	**1**

CONCLUSION

The random forest is suitable for this dataset as it consists of multiple decision trees. Based on learners' categories system can provide assistance to teachers to provide the best learning style.

REFERENCES

[1] https://www.analyticsvidhya.com/blog/2020/05/decision-tree-vs-random-forest- algorithm/

[2] https://www.sciencedirect.com/science/article/abs/pii/S0048969718324653

[3] https://www.upgrad.com/blog/random-forest-vs-decision-tree/#:~:text=A%20decision%20tree%20combines%20some,forest%20model%20needs%20rigorous%20training.