

# Prediction of Study Track using Decision Tree and Aptitude Test

Deepali Joshi  
Assistant Professor  
Mumbai, India

## ABSTRACT

In order to succeed in the competitive environment it is essential to be successful in academics. The basic education is from 1st to 10th standard and once 10th standard is complete there are various courses that can be selected by the students such as Science, Commerce, Arts and Diploma. It becomes difficult to identify the suitable stream. The proposed system can be used to solve the problem. The proposed system implements two methods to generate solution. The Aptitude Test predicts the suitable stream depending upon the intellectual capability of the student. Apart from this prediction is done depending on the ssc marks obtained by the student. The combination of these two methods is also implemented to get more accurate results. The proposed system not only predicts the stream but also specifies the colleges for the predicted stream depending upon the location of student and also gives information about vocational courses that the student can pursue after 10th.

## Keywords

Data Mining, C4.5, NBTree, ssc, ssc-marks, Precision, Recall, fMeasure

## 1. INTRODUCTION

As the basic education is complete the student have to select the stream and it is essential to select the appropriate stream. If correct field is selected by students then they will be successful in their careers [1]. Once ssc is complete various courses are available like Science, Commerce, Arts and Diploma [6]. Due to various available options it becomes difficult to choose the suitable field. A technique is required through which the students will be able to evaluate the suitable stream. Solutions are available in order to solve this problem but they do not provide appropriate results. One method which is used to specify the stream is the Aptitude Test Method. The aptitude test consists of Questions and Answers. The student have to answer the given question. The questions are based on streams like Science, Commerce, Arts and Diploma. The aptitude test intellectual capability of the student can be assessed and the suitable stream can be predicted to the student. The paper contains detailed information about Building the Model, Data Collection, Tools, Implementation, Solution, Results and Conclusion and followed by references.

## 2. BUILDING THE MODEL

Data Mining means extracting the knowledge from large amount of data. Decision Tree is a classification technique[9], it is a tree structure which consists of nodes, leaf node, non-leaf node, root node. Various algorithms are available for generating decision tree. The algorithms used in the system are C4.5, NBTree. The C4.5 algorithm is used for predicting the stream based on ssc marks. The NBTree algorithm will predict the stream based on aptitude test.

In order to generate the results WEKA API is used. The data has to be supplied to Weka API and is supplied in the form of arff files. Three arff files are used Train, Test, New. The train data is supplied to the WEKA API so that it can learn from the trained data set and can generate the results. The generated result will again be stored in the training data set so it can be used for the next input.

## 3. DATA COLLECTION

In order to implement the system data is required and this data is to be provided to the algorithms to generate the results. To predict the streams based on ssc marks SSC mark-sheets were gathered from the students. The SSC marks were also gathered from college students.

To predict the streams based on aptitude test the questions and their correct answer along with three other answer options will be displayed. The data for aptitude test is gathered from the 10th standard text books as well as the internet. The questions and answers are from streams like Science, Commerce, Arts, Diploma.

## 4. RELATED WORK

### 4.1 C4.5

The C4.5 algorithm is a successor to ID3 and it can handle categorical as well as continuous attributes to build a decision tree[6]. The C4.5 algorithm uses the divide and conquer technique. The C4.5 is used due to various advantages, It can handle the continuous and discrete values, It can handle missing values. In order to handle continuous attributes, C4.5 splits the attribute values into two partitions based on the selected threshold such that all the values above the threshold as one child and the remaining as another child. It also handles missing attribute values.

### 4.2 NBTree

NB Tree is a hybrid of Naive Bayes classifier and classification[6]. The NB Tree model can be described as a decision tree with nodes and branches. Given the A set of instances of the algorithm evaluation node "Practice" for each division of the property.

## 5. DATASET

The data set used in this research was obtained by collecting the ssc mark-sheets from students as well as the stream selected by them. As the students complete ssc they can select courses such as Arts, Science, Commerce, Diploma. There are four classes based on which the algorithms give the prediction.

Table 1: Prediction Classes

Class
Arts
Science

Commerce
Diploma

## 6. EVALUATION

### 6.1 Precision

Precision is a measure of the accuracy provided that a specific class has been predicted. It is defined by:  $Precision = tp / (tp + fp)$ , where tp and fp are the numbers of true positive and false positive predictions for the considered class.

The Precision for four classes can be calculated as:

$$1. \text{ PrecisionDiploma} = \frac{tpDiploma}{tpDiploma + eba + eCA + eDA} \quad (1)$$

$$2. \text{ PrecisionCommerce} = \frac{tpCommerce}{tpCommerce + eAB + eCB + eDB} \quad (2)$$

$$3. \text{ Precision Science} = \frac{tpScience}{tpScience + eAC + eBC + eDC} \quad (3)$$

$$4. \text{ Precision Arts} = \frac{tpArts}{tpArts + eAD + eBD + eCD} \quad (4)$$

### 6.2 Recall

Recall is a measure of the ability of a prediction model to select instances of a certain class from a data set. It is commonly also called sensitivity, and corresponds to the true positive rate. It is defined by the formula:

$$\text{Recall} = \text{Sensitivity} = \frac{tp}{tp + fn}$$

where tp and fn are the numbers of true positive and false negative predictions for the considered class. tp + fn is the total number of test examples of the considered class.

The Recall for four classes can be calculated as:

$$1. \text{ RecallDiploma} = \frac{tpDiploma}{tpDiploma + eAB + eAC} \quad (5)$$

$$2. \text{ Recall Commerce} = \frac{tpCommerce}{tpCommerce + eBA + eBC + eBD} \quad (6)$$

$$3. \text{ Recall Science} = \frac{tpScience}{tpScience + eCA + eCB + eCD} \quad (7)$$

$$4. \text{ Recall Arts} = \frac{tpArts}{tpArts + eDA + eDB + eDC} \quad (8)$$

### 6.3 fMeasure

fMeasure is a measure that combines precision and recall is the harmonic mean of precision and recall.

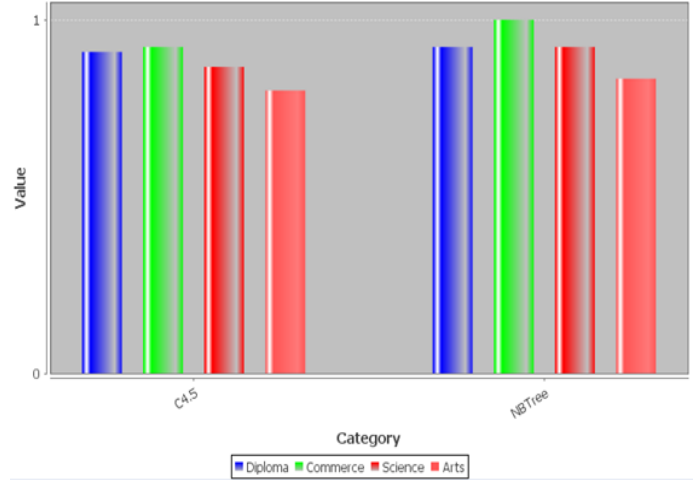
$$\bullet \text{ F-measure} = \frac{2 * (\text{Precision} * \text{Recall})}{(\text{Precision} + \text{Recall})} \quad (9)$$

## 7. RESULTS

The Table represent the results for C4.5 and NBTree algorithms based on four classes Arts, Science, Commerce, Diploma. For students classified as Arts class the prediction value is most accurate given the recall value =1, precision value=0.833 and fmeasure=0.909. For students classified as Commerce class the prediction value is most accurate given the recall value =0.923, precision value=1 and fmeasure=0.96. For students classified as Diploma class the prediction value is most accurate given the recall value =0.923, precision value=0.923 and fmeasure=0.923. For students classified as Science class the prediction value is not accurate given the recall value =0.923, precision value=0.923 and fmeasure=0.923.

**Table 2: Precision Statistics Comparison**

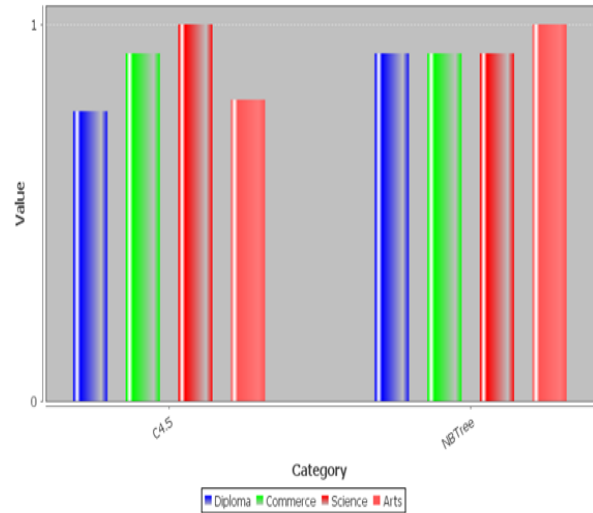
Class	C4.5	NBTree
Arts	0.8	<b>0.833</b>
Science	0.867	<b>0.923</b>
Commerce	0.923	<b>1</b>
Diploma	0.909	<b>0.923</b>



**Fig 1: Precision Statistics**

**Table 3: Recall Statistics Comparison**

Class	C4.5	NBTree
Arts	0.8	<b>1</b>
Science	1	0.923
Commerce	0.923	0.923
Diploma	0.769	<b>0.923</b>



**Fig 2: Recall Statistics**

**Table 4: fMeasure Statistics Comparison**

Class	C4.5	NBTree
Arts	0.8	<b>0.909</b>
Science	0.929	0.923
Commerce	0.923	<b>0.96</b>
Diploma	0.833	<b>0.923</b>

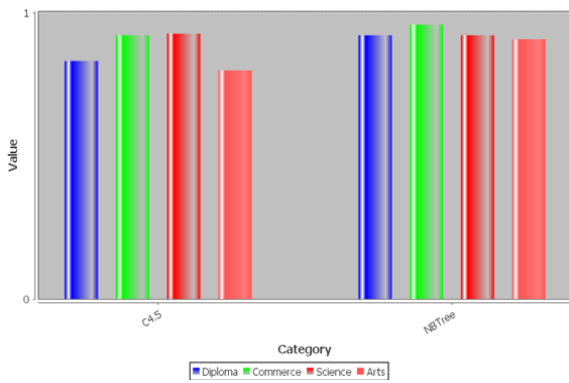


Fig 3: fMeasure Statistics

The Aptitude Test method was conducted for twenty-five students of Abhinandan Classes. Each and every student personally gave the aptitude test on separate machines and the questions displayed to them were different to make sure that their intellectual capability is tested in a efficient way. The aptitude test method worked successfully and predicted the suitable stream for individual student.

The chart specifies the result of the Aptitude Test. There are four streams Science, Commerce, Arts, Diploma. Out of twenty-five students, science was the suitable stream for eight students and in percentage it can be specified as 32%, commerce for five students and percentage is 20%, arts for three students and percentage is 12%, diploma for nine students and can be specified as 20%.

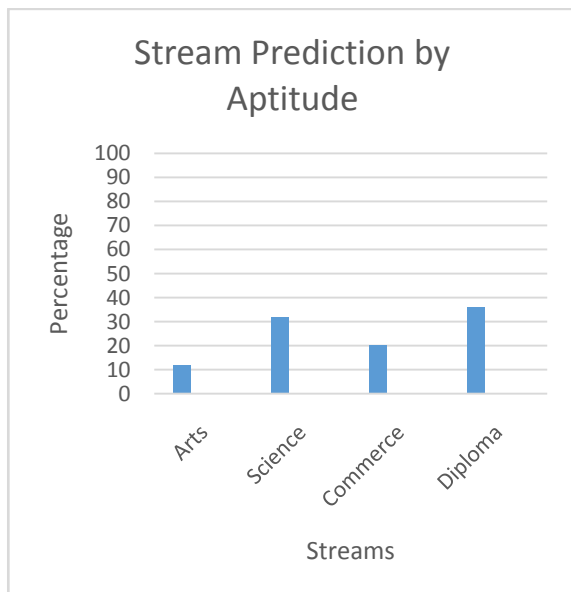


Fig 4: Prediction Statistics

## 8. CONCLUSION

The system takes input from the students and predicts the stream suitable for them due to which the student will be successful in the academics. The system predicts the result on the basis of ssc marks and the aptitude test. The prediction by ssc marks also considers marks of combination of subjects to give the prediction. Based on the location of student the colleges providing the predicted stream will be specified. The detailed information about the colleges is provided to the students as well as the courses offered by them, so that the searching time can be minimized. Apart from the information

about various vocational courses is given to the student. The system uses two algorithms for prediction C4.5 and NBTree. There are four classes Arts, Science, Commerce, Diploma. The algorithms are compared using several factors such as precision, recall, fmeasure and the analysis shows that prediction by NBTree is accurate. The NBTree algorithm gives the most accurate for Arts, Commerce, Diploma classes. The future scope is that it can be implemented for 12<sup>th</sup> standard students also to help them to select the suitable streams.

## 9. REFERENCES

- [1] Ahmad I., Manarvi, I., Ashraf, N. "Predicting university performance in a subject based on high school majors", 978-1-4244-4136-5/09/ ©2009 IEEE
- [2] Zhiwu Liu, Xiuzhi Zhang. "Prediction and Analysis for Students' Marks Based on Decision Tree Algorithm", Intelligent Networks and Intelligent Systems (ICINIS), 2010 3rd International Conference on Digital Object Identifier:10.1109/ICINIS.2010.59. Publication Year: 2010, Page(s): 338 – 341
- [3] Anupama Kumar S, Vijayalakshmi M.N., "Mining of student academic evaluation records in higher education", International Conference on Recent Advances in Computing and Software Systems (RACSS), 2012 Digital Object Identifier:10.1109/RACSS.2012.6212699, Publication Year: 2012 IEEE , Page(s): 67 – 70
- [4] Bunkar, K, Singh U.K., Pandya B, Bunkar R, "Data mining: Prediction for performance improvement of graduate students using classification", WOCN, 2012 Digital Object Identifier: 10.1109/WOCN.2012.6335530 Publication Year: 2012 IEEE, Page(s): 1 – 5
- [5] Garcia, E.P.I. ; Mora, P.M., "Model Prediction of Academic Performance for First Year Students", Artificial Intelligence (MICAI), 2011 10th Mexican International Conference on Digital Object Identifier: 10.1109/MICAI.2011.28 Publication Year: 2011 IEEE , Page(s): 169 – 174
- [6] Pumpuang, P., Srivihok, A., Praneetpolgrang, "Comparisons of classifier algorithms: Bayesian network, C4.5, decision forest and NBTree for Course Registration Planning model of undergraduate students", Systems, Man and Cybernetics, 2008. SMC 2008. IEEE International Conference on Digital Object Identifier: 10.1109/ICSMC.2008.4811865 Publication Year: 2008 IEEE, Page(s): 3647 – 3651
- [7] Qasem A. Al-Radaideh, Ahmad Al Ananbeh, and Emad M. Al-Shawakfa A "Classification Model For Predicting The Suitable Study Track For School Students". IJRRAS 8 (2) August 2011
- [8] Sunita Beniwal, jitender Arora "Classification and Feature Selection Techniques in Data Mining", IJERT AUGUST 2012
- [9] Heena Sharma, Navdeep Kaur Kaler "Data Mining with Improved and Efficient Mechanism in Clustering Analysis and Decision Tree as a Hybrid Approach" (IJITEE) ISSN: 2278-3075, Volume-2, Issue-5, April 2013
- [10] Vaibhav P. Vasani, Rajendra D. Gawali "Classification and performance evaluation using data mining

- algorithms”, IJRSET ISSN: 2319-8753, Vol. 3, Issue 3, March 2014
- [11] Mr. Shridhar Kamble, Mr. Aaditya Desai “Evaluation and Performance Analysis of Machine Learning Algorithms”, IJESRT ISSN: 2277-9655, May, 2014
- [12] Mohd. Mahmood Ali<sup>1</sup>, Mohd. S. Qaseem<sup>2</sup>, Lakshmi Rajamani<sup>3</sup>, A. Govardhan<sup>4</sup>, “EXTRACTING USEFUL RULES THROUGH IMPROVED DECISION TREE INDUCTION USING INFORMATION ENTROPY”, International Journal of Information Sciences and Techniques (IJIST) Vol.3, No.1, January 2013
- [13] Abeer Badr El Din Ahmed, Ibrahim Sayed Elaraby, “Data Mining: A prediction for Student's Performance Using Classification Method “World Journal of Computer Application and Technology 2(2): 43-47, 2014, DOI: 10.13189/wjcat.2014.020203
- [14] Dorina Kabakchieva, “Predicting Student Performance by Using Data Mining Methods for Classification”, CYBERNETICS AND INFORMATION TECHNOLOGIES • Volume 13, No 1 , ISSN: 1311-9702; Online ISSN: 1314-4081
- [15] Ajay Kumar Pal , Saurabh Pal , “Classification Model of Prediction for Placement of Students”, IJ.Modern Education and Computer Science, 2013, 11, 49-56 Published Online November 2013 in MECS (<http://www.mecs-press.org/>)DOI: 10.5815/ijmecs.2013.11.07
- [16] M. Hanumanthappa, M Narayana Swamy, “Predicting Academic Success from Student Enrolment Data using Decision Tree Technique”, International Journal of Applied Information Systems (IJ AIS) – ISSN : 2249-0868 Foundation of Computer Science FCS, New York, USA Volume 4– No.3, September 2012 – [www.ijais.org](http://www.ijais.org)
- [17] Sunita B Aher, Mr. LOBO L.M.R.J, “Data Mining in Educational System using WEKA”, International Conference on Emerging Technology Trends (ICETT) 2011 Proceedings published by International Journal of Computer Applications® (IJCA)
- [18] M.I. López, J.M Luna, C. Romero, S. Ventura, “Classification via clustering for predicting final marks based on student participation in forums”
- [19] Neelamadhab Padhy <sup>1</sup>, Dr. Pragnyaban Mishra <sup>2</sup>, and Rasmita Panigrahi <sup>3</sup>, “The Survey of Data Mining Applications And Feature Scope”, International Journal of Computer Science, Engineering and Information Technology (IJCEIT), Vol.2, No.3, June 2012
- [20] Surjeet Kumar Yadav, Saurabh Pal, “Data Mining: A Prediction for Performance Improvement of Engineering Students using Classification”, World of Computer Science and Information Technology Journal (WCSIT) ISSN: 2221-0741 Vol. 2, No. 2, 51-56, 2012
- [21] Surjeet Kumar Yadav <sup>1</sup>, Brijesh Bharadwaj<sup>2</sup>, Saurabh Pal<sup>3</sup>, “Data Mining Applications: A comparative Study for Predicting Student’s performance”, INTERNATIONAL JOURNAL OF INNOVATIVE TECHNOLOGY & CREATIVE ENGINEERING (ISSN:2045-711) VOL.1 NO.12 DECEMBER
- [22] J. Han and M. Kamber, “Data Mining: Concepts and Techniques”, Morgan Kaufmann, 2nd, 2006
- [23] T.Miranda Lakshmi, A.Martin, R.Mumtaj Begum, Dr.V.Prasanna Venkatesan, “An Analysis on Performance of Decision Tree Algorithms using Student’s Qualitative Data”, IJ.Modern Education and Computer Science, 2013, 5, 18-27 Published Online June 2013 in MECS (<http://www.mecs-press.org/>),DOI: 10.5815/ijmecs.2013.05.03
- [24] Elakia, Gayathri, Aarthi, Naren J, “Application of Data Mining in Educational Database for Predicting Behavioural Patterns of the Students”, Elakia et al, / (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 5 (3) , 2014, 4649-4652
- [25] Pimpa Cheewaparakobkit, “Study of Factors Analysis Affecting Academic Achievement of Undergraduate Students in International Program”, Proceedings of the International MultiConference of Engineers and Computer Scientists 2013 Vol I, IMECS 2013, March 13 - 15, 2013, Hong Kong