Classification Technique for Predicting Learning Behavior of Student in Higher Education

Mrs. Varsha. P. Desai

Assistant Professor, V. P. Institute of Management Studies & Research, Sangli, Maharashtra, India
Affiliated to Shivaji University, Kolhapur, Maharashtra, India

ABSTRACT

In education system it is very important to decide learning behavior of students. Today there is huge competition in higher educational institutes. Quality education is essential for facing new educational challenges. Educational Data Mining is useful to classify students according to their knowledge and learning behavior. It helps teachers to implement different teaching methodology as per learning behavior of student. Researcher used Naïve Bayes classification technique on training data set of students. Classification is a supervised learning approach which categorized data into predefined classes. The implementation is carried out using C#. Algorithm is implemented on set of multivalued attributes to predict slow learner, average learner and fast learner students. The objective of researcher is to extract hidden knowledge from dataset for prediction of learning behavior of student.

KEYWORD: *Training Dataset, Supervised, Unsupervised, Machine learning, Data Mining.*

I. INTRODUCTION

Data Mining is a process of discovering knowledge from database. It is a technique to identify patterns and determine relationship between objects in dataset. Data mining motivates various applications in machine learning to learn from data. It consists of many algorithms which are based on supervised and unsupervised learning. There are different techniques of data mining like classification, clustering, predictive analysis, association rule mining, sequence mining, graph mining, regression and time series analysis etc. Selection and implementation of best

suitable algorithm for getting optimum solution to the problem is a challenging task in data mining.

Data mining plays vital role in education system. Predicting learning behavior of student is very critical process. Learning behavior of student depend of different factors like gender, family background, location, age, interest, strength, weakness, culture, curriculum etc. Today education system creates tremendous carrier opportunities in the front of students. It is challenging work for teacher to provide education as per student need and interest. Learning student behavior is very essential for getting better teaching outcome as well as student's satisfaction. A Classification technique in data mining helps teachers to predict student behavior and selecting appropriate teaching methodology to enhance teaching and learning process.

II. Literature Review:

Researcher has gone through previous research related to classification techniques in data mining. It is observed that, Naïve Bayes classification algorithm is used for student's performance classification. Web mining and multifactor analysis technique is implemented for prediction [3]. Decision tree, Random forest and Naïve Bayes theorem is used for classification of student behavior. Researcher evaluate results of all three algorithms and it is found that Naïve Bayes method gives better results than other classification techniques.[4] Naïve Bays algorithm is implemented for slow Lerner prediction using python and accuracy is compared using WEKA data mining tool.

According to literature review it is found that Naïve Bayes is suitable classification algorithm for multi attribute analysis. It is essential to develop user friendly application which useful in any education sector. Researcher developed application using C# for predicting learning behavior of student by implementing Naïve Bayes theorem.

III. Classification Techniques:

Classification is a supervised learning method where data is divided into different categories or classes. The objective of classification to predict target class for given dataset. There are various techniques of classification like decision tree, Naïve Bayes classifier, nearest neighbor approach, artificial neural network these are important techniques of classification. Accuracy of target prediction is depends upon selection of classification technique. In many real life situations classification is fundamentally probabilistic, it is uncertain to which class record is belong.[1]

IV. Naïve Bayes Classifier:

Bayesian classification is based on Bayes theorem. The posterior probability of the class that a record belongs to is an approximated using prior probability which drawn from training dataset. Classification model estimate the likelihood of the record belonging

to each class. The class with highest prevents for Y to happen when events for X probability becomes the class label for the record.[2]

Definition of Bayes Theorem: Given two random variables X and Y, each of them taking a specific value corresponds to a random event. A conditional probability P(X/Y) represents the probability of events for Y to happen when event for X have already occurred.[2]

$$P(X/Y) = \frac{P(X/Y).P(Y)}{P(X)}$$

$$P(Y/X) = \frac{P(X/Y).P(Y)}{P(Y)}$$

V. Training Dataset:

Following table shows training dataset of MCA I year student dataset. Here researcher is interested to predict learning behavior of student from given training dataset using Naïve Bayes algorithm. Student data consists of different attributes like Gender, Area, SSC_Medium, SSC_Percentage, HSC_faculty, Math_At_HSC,Graduation_Marks,Admission_Type, Entrance_Rank,ParentsIncome,,Attendance,Communication_Skill, Learning_Behavior (Class Label) etc.

Table 1: Training Dataset:

Sr. No	Gender	Area	SSC_Medium	SSC_Percentage	HSC_Faculty	HSC_Percentage	Maths At HSC
1	M	Rural	English	Excellent	Commerce	Poor	Yes
2	M	Urban	English	Good	Science	Good	Yes
3	M	Urban	English	Good	Commerce	Poor	No
4	F	Urban	Marathi	Poor	Arts	Good	Yes
5	M	Rural	Marathi	Poor	Science	Excellent	No
6	M	Rural	Marathi	Average	Commerce	Poor	No
7	F	Urban	Marathi	Excellent	Commerce	Excellent	Yes
8	F	Rural	Marathi	Poor	Commerce	Poor	No
9	M	Rural	Marathi	Excellent	Science	Poor	No
10	F	Urban	English	Poor	Science	Good	Yes

Graduation Marks	Admission Type	Entrance Rank	Parents Income	Attendance	Communication Skill	Learning Behavior
Excellent	MC	Good	High	Poor	Good	Slow
Poor	ER	Poor	Medium	Average	Poor	Fast
Good	MC	Good	Low	Good	Good	Average
Good	MC	Average	Low	Good	Good	Slow
Poor	MC	Poor	High	Average	Poor	Fast
Excellent	ER	Good	Medium	Poor	Excellent	Average
Poor	ER	Good	Medium	Average	Poor	Slow
Good	ER	Average	Low	Average	Excellent	Fast
Good	ER	Good	Low	Good	Good	Fast
Poor	ER	Good	High	Average	Excellent	Average

VI. Student related Variables:

Attributes	Possible Values
Gender	M,F
Area	Urban, Rural
SSC_Medium	English, Marathi
SSC_Percentage	>=70 :Excellent, >=60 &<70:Good, >=50&<60:Average, <50:Poor
HSC_Faculty	Commerce, Arts, Science
HSC_Percentage	>=70 :Excellent, >=60 &<70:Good, >=50&<60:Average, <50:Poor
Maths At HSC	Yes, No
Graduation Marks	>=70 :Excellent, >=60 &<70:Good, >=50&<60:Average, <50:Poor
Admission Type	MC: Management Cota, ER- Entrance Round
Entrance Rank	Good, Average, Poor
Parents Income	>=10 Lacs: High, >=5Lacs & <10Lacs: Medium, <=5 Lacs: Low
Attendance	Below 50: low, >50 &<70: Medium, >70: High
Communication Skill	Good, Poor, Excellent
Learning Behavior	Slow, Fast, Average (Class Labels)

VII. Data Pre-processing:

Data was pre-processed by performing following operations [3]:

1. Converting all fields to categories.
2. Features combine to reduce dimensionality.
3. Missing values are replaced by frequently occurring values.

VIII. Algorithm:

1. Import dataset into Sqlserver
2. Find probability of each class.
3. Select parameter set as per input requirement.
4. For each input record:
i. For each attribute:
   A. Entities are divided into different categories according to categorical data.
   B. Probability is calculated from training dataset.
5. For each attribute in testing dataset
i. For each attribute:
   A. Calculate probability and classify the data accordingly
   B. Return the diagnosis parameter and calculated probability of each class [4].

C. Compare class wise probability value and Return final classification which has highest probability.

IX. Implementation of algorithm:

Here Naïve Bayes algorithm is implemented on above dataset. C# is used for stepwise implementation of algorithm and predicting data for unknown tuple/record.

Algorithm is implemented to predict learning behavior of student with following known attribute values:

X= Gender=M, Area=Rural, SSC_Medium=English, SSC_Percentage=Poor, HSC_Faculty=Commerce, HSC_percentage=Good, Maths_At_HSC=Yes, Graduation_Marks:Poor, Admission_Type=MC, Entrance_Rank=Good, parents_Income=Low, Attendance=Average, Communicaton_Skill=Good.

In above problem there are three classes:

C1: Learning Behavior Slow

C2: Learning Behavior Fast,

C3: Learning Behavior Average.

Here we need to predict whether X belongs to which class.

$P(X/C1)=0.33*0.33*0.33*0.33*0.66*0.33*1*0.33*0.66*0.66*0.33*0.33*0.66=2.66$

$P(X/C2)=0.66*0.33*0.66*0.33*0.66*0.33*0.33*0.33*0.33*1*0.33*0.33*0.33=1.33$

$P(X/C3)=0.75*0.75*0.25*0.5*0.25*0.25*0.25*0.5*0.25*0.25*0.5*0.75*0.25=3.21$

$P(X/C1)*P(C1)=2.66*0.3=0.798$

$P(X/C2)*P(C2)=1.33*0.3=0.399$

$P(X/C3)*P(C3)=3.21*0.4=1.284$

$P(X/C3)*P(C3)$ gives highest probability so X belongs to class C3.

According to Naïve Bayes theorem it is predicted that given tuple X belongs to class C3. Which means that there is highest probability that student is Fast Lerner.

X. Finding:

Implementation of Naïve Bayes theorem using C# we can find out Fast, Slow and Average learners.

Conclusion:

Naïve bays theorem is implemented using C# to determine Slow Learner, Average Lerner and Fast Learner. This application is useful in education system to categories student according to their learning behavior. Proposed application is very user friendly and applicable for any higher education sector. It helps teachers to implement different teaching and learning techniques for providing quality education to the students. Successful implementation of this model will improve overall result and learning interest among students.

REFERENCES:

1. Jiawei Han and Micheline Kamber, *Data Mining Concepts and Techniques* ISBN-978-81-312-0535-8.

2. Hongbo Du, *Data Mining Technique,* ISBN-978-81-315-1955-4.

3. K. Prasada Rao, M. V. P. Chandra Sekhara Rao, B. Ramesh, *Predicting Learning Behavior of Students using Classification Techniques, International Journal of Computer Applications (0975 – 8887) Volume 139 – No.7, April 2016.*

4. Swati and Rajinder Kaur, *Using Factor Classification for the Slow Learner Prediction over Various Class of Student Dataset,* **Indian Journal of Science and Technology,** *Vol 9(48), DOI: 10.17485/ijst/2016/v9i48/103651, December 2016, ISSN (Online): 0974-564.*

5. Swati and Rajinder Kaur, *Multifactor Naïve Bayes Classification For The Slow Learner Prediction Over Multicass Student Dataset,* International Journal on Computational Science & Applications (IJCSA) Vol.6, No. 4, August 2016

6. Shiwani Rana*, Roopali Garg, *Slow Learner Prediction using Multi-Variate Naïve Bayes Classification Algorithm,* Department of Information Technology, UIET, Panjab University, Chandigarh, India. 02 December 2016.

7. R. Kohavi, "*Scaling up the accuracy of Naïve Bayes classifiers: a decision-tree hybrid,*" Proc. International Conference on Knowledge Discovery and Data Mining (KDD 96), ACM, Aug. 1996, pp. 202-207.

8. C. G. Nespereira, E. Elhariri, N. El-Bendary, A. F. Vilas, and R. P. D. Redondo, *"Machine learning based classification approach for predicting student's performance in blended learning,"* Proc. International Conference on Advanced Intelligent System and Informatics (AISI 15), Springer, Nov. 2015, pp. 47-56.

9. Sudha M, Kumaravel A. *Performance comparison based on attribute selection tools for data mining.* Indian Journal of Science and Technology. 2014 Nov; 7(S7):1–5.

10. Weka, University of Waikato, New Zealand, http://www.cs.waikato.ac.nz/ml/weka/