International Journal of Trend in Scientific Research and Development (IJTSRD)
Volume 3 Issue 5, August 2019 Available Online: www.ijtsrd.com e-ISSN: 2456 – 6470

Using Naive Bayesian Classifier for Predicting Performance of a Student

Khin Khin Lay, Aung Cho

University of Computer Studies, Maubin, Myanmar

How to cite this paper: Khin Khin Lay | Aung Cho "Using Naive Bayesian Classifier for Predicting Performance of a Student" Published in International Journal of Trend in Scientific Research and Development (ijtsrd), ISSN: 2456-6470, Volume-3 | Issue-5, August 2019, pp.1387-1391, **https://doi.org/10.31142/ijtsrd26463**

IJTSRD26463

ABSTRACT
Data mining techniques play an important role in data analysis. For the construction of a classification model which could predict performance of students, particularly for engineering branches, a decision tree algorithm associated with the data mining techniques have been used in the research. A number of factors may affect the performance of students. Data mining technology which can relate to this student grade well and we also used classification algorithms prediction. We proposed student data classification using Naive Bayesian Classifier.

KEYWORDS: *Classification, Naive Bayesian, Data Mining, Predicting Performance*

I. INTRODUCTION
Data Mining Advancement in technology has brought great growth in the volume of the data available on the internet, digital libraries, news sources and company-wide intranets. This makes a huge number of databases and information repositories available, but it is impossible to manually organize, analyze and retrieve information from this data.

This generates an eminent need of methods that can help users to efficiently navigate, summarize, and organize the data so that it can further be used for applications ranging from market analysis, fraud detection and customer retention etc. Therefore, the techniques that perform data analysis and may uncover important data patterns are needed. One of these techniques is known as data mining.

There are increasing research interests in using data mining in education. This new emerging field, called Educational Data Mining, concerns with developing methods that discover knowledge from data originating from educational environments [4]. Educational Data Mining uses many techniques such as Decision Trees, Neural Networks, Naïve Bays, K- Nearest neighbor, and many others.

The main aim of this paper is to use data mining methodologies to study students' performance in the courses. Data mining provides many tasks that could be used to study the student performance. Here the classification tasks is used to evaluate student's performance and as there are many approaches that are used for data classification, the decision tree and Naive Bays method are used [4]. Decision trees can easily be converted to classification rules Decision tree algorithms, such as ID3 (Iterative Dichotomies), C4.5, and CART (Classification and Regression Trees) [3]. Naïve Bayesian classifiers assume that the effect of an attribute value on a given class is independent of the values of the other attribute [3].This paper explores the accuracy of Decision tree and Naive Bays techniques for predicting student performance.

II. RELATED WORK
In order to predict the performance of students the searcher took into consideration the work of other 14 A Decision Tree Approach forPredicting Students Academic Performance researchers that are in the same direction. Other researchers have looked at the work of predicting students' performance by applying many approaches and coming up with diverse results.

Classification is a classic data mining technique based on machine learning. Basically classification is used to classify Each item in a set of data into one of predefined set of classes or groups. Classification method makes use of mathematical Techniques such as decision trees, linear programming, neural network and statistics. In classification, once the software is made that can learn how to classify the data items into groups.

Three supervised data mining algorithms, i.e. Bayesian, Decision trees and Neural Networks which were applied by [1] on the preoperative assessment data to predict success in a course (to produce result as either passed or failed) and the performance of the learning methods were evaluated based on their predictive accuracy, ease of learning and user friendly characteristics. The researchers observed that that this methodology can be used to help students and teachers to improve student's performance; reduce failing ratio by taking appropriate steps at right time to improve the quality of learning.

The authors in [9] analyzed student's performance data using classification algorithm named ID3 to predict student's marks at the end of the semester. This was applied for master of computer applications course from 2007 to 2010 in VBS Purvanchal University, Jaunpur. Their study aimed to help students and teachers find ways to improve students'

performance. Data was collected from 50 students, and then a set of rules was extracted for their analysis. Another study that focused on the behavior to improve students' performance using data mining techniques is illustrated in [10]. The data consisted of 151 instances from a data base management system course held at the Islamic University of Gaza. The data was collected from personal records and academic records of students. The author performed the data mining techniques, namely: association rules, classification, and clustering and outlier detection. The results revealed useful information from association rules and classification models.

III. OUR PROPOSED METHOD

a. Naïve Bayesian Classifier

Student performance is predicted using a data mining technique called classification rules. The naïve Bayes classification algorithm is used by the administrator to predict student performance based on performance detail. The algorithm is a simple probabilistic classifier that calculates a set of probabilities by counting the frequency and combinations of values in a given dataset. The algorithm uses the Bayes theorem and assumes that all attributes are independent given the value of the class variable. This conditional independence assumption rarely holds true in real-world applications, hence the characterization as naïve. However, the algorithm tends to perform well and learn rapidly in various supervised classification problems [11].

The naïve Bayesian classifier works as follows:

1. We let T be a training set of samples, each with their class labels. k classes, C1,C2, . . . ,Ck exist. Each sample is represented by an n-dimensional vector, X = {x1, x2, . . . , xn} that depicts n measured values of the n attributes, A1,A2, . . . , An, respectively.

2. Given a sample X, the classifier will predict that X belongs to the class having the highest a posteriori probability,

Conditioned on X. X is predicted to belong to the class Ci if and only if

P(Ci|X) > P(Cj |X) for 1<= j <= m, j ≠i.

Thus, we find the class that maximizes P(Ci|X). The class Ci for which P(Ci|X) is maximized is called the maximum posteriori hypothesis. By Bayes' theorem,

P(Ci|X) = P(X|Ci) P(Ci) P(X)

3. Given that P(X) is the same for all classes, only P(X|Ci)P(Ci) needs to be maximized. If the class a priori probabilities, P(Ci), are not known, then we assume that the classes are equally likely, that is, P(C1) = P(C2) = . . . = P(Ck), and we would therefore maximize P(X|Ci). Otherwise, we maximize P(X|Ci)P(Ci). The class a priori probabilities may be estimated by P(Ci) = freq(Ci, T)/|T|.

4. Given data sets with many attributes, computing P(X|Ci) would be computationally expensive. To reduce computation in evaluating P(X|Ci)P(Ci), the naïve assumption of class conditional independence is made. This assumption presumes that the values of the attributes are conditionally independent of one another, given the class label of the sample.

$$P(X|Ci) \approx \prod_{k=1}^{n} P(Xk|Ci) \qquad \text{Eq(1.1)}$$

The probabilities P(x1|Ci), P(x2|Ci), . . . , P(xn|Ci) can easily be estimated from the training set. Here, xk refers to the value of attribute Ak for sample X.

5. To predict the class label of X, P(X|Ci)P(Ci) is evaluated for each class Ci. The classifier predicts that the class label of X is Ci if and only if it is the class that maximizes P(X|Ci)P(Ci) [12].

IV. Training Dataset

The first step in this paper is to collect data. It is important to select the most suitable attributes which influence the student performance. We have training set of 30 under graduate students. We were provided with a training dataset consisting of information about students admitted to the first year in Table I.

TABLE.I Training dataset

Sr. no.	Roll no.	Attend-ance	Apti- tude	Assign-ment	Test	Present-ation	Grade
1	IT1	Good	Avg	Yes	Pass	Good	Excellent
2	IT2	Good	Avg	Yes	Pass	Good	Excellent
3	IT 3	Good	Avg	Yes	Pass	Good	Excellent
4	IT4	Good	Avg	Yes	Pass	Good	Excellent
5	IT5	Good	Avg	Yes	Pass	Good	Excellent
6	IT6	Avg	Avg	Yes	Pass	Avg	Good
7	IT7	Poor	Good	Yes	Pass	Avg	Good
8	IT8	Avg	Good	Yes	Pass	Avg	Good
9	IT9	Avg	Good	Yes	Pass	Avg	Good
10	IT10	Poor	Poor	No	Fail	Poor	Fail
11	IT11	Poor	Poor	No	Fail	Poor	Fail
12	IT12	Avg	Age	Yes	Pass	Age	Excellent
13	IT13	Good	Good	Yes	Pass	Good	Excellent
14	IT14	Good	Good	Yes	Pass	Good	Excellent
15	IT15	Good	Good	Yes	Pass	Good	Excellent
16	IT16	Good	Good	Yes	Pass	Good	Excellent
17	IT17	Good	Avg	Yes	Pass	Good	Excellent
18	IT18	Good	Avg	Yes	Pass	Good	Excellent
19	IT19	Good	Avg	Yes	Pass	Good	Excellent
20	IT20	Good	Poor	Yes	Pass	Good	Excellent

21	IT21	Good	Poor	Yes	Pass	Good	Excellent
22	IT22	Good	Poor	Yes	Pass	Good	Excellent
23	IT23	Good	Poor	Yes	Pass	Good	Excellent
24	IT24	Good	Poor	Yes	Pass	Good	Excellent
25	IT25	Poor	Poor	No	Fail	Poor	Fail
26	IT26	Avg	Good	Yes	Pass	Avg	Good
27	IT27	Poor	Good	No	Fail	Poor	Fail
28	IT28	Good	Good	Yes	Pass	Good	Excellent
29	IT29	Good	Good	Yes	Pass	Good	Excellent
30	IT30	Good	Good	Yes	Pass	Good	Excellent

A. Classification Steps

TABLE.II Frequency Tables

Frequency Table		Grade		
		Excellent	Good	Fail
Attendance	Good	20	0	0
	Average	1	4	0
	Poor	0	1	4

Frequency Table		Grade		
		Excellent	Good	Fail
Aptitude	Good	7	4	1
	Average	9	1	0
	Poor	5	0	3

Frequency Table		Grade		
		Excellent	Good	Fail
Assignment	Yes	21	5	0
	No	0	0	4

Frequency Table		Grade		
		Excellent	Good	Fail
Test	Pass	20	5	0
	Fail	0	0	4

Frequency Table		Grade		
		Excellent	Good	Fail
Presentation	Good	20	0	0
	Average	1	5	0
	Poor	0	0	4

TABLE.III Likelihood Tables

Likelihood Table		Grade			
		Excellent	Good	Fail	
Attendance	Good	20/21	0/5	0/4	20/30
	Average	1/21	4/5	0/4	5/30
	Poor	0/21	1/5	4/4	5/30
		21/30	5/30	4/30	

Likelihood Table		Grade			
		Excellent	Good	Fail	
Aptitude	Good	7/21	4/5	1/4	12/30
	Average	9/21	1/5	0/4	10/30
	Poor	5/21	0/5	3/4	8/30
		21/30	5/30	4/30	

Likelihood Table		Grade			
		Excellent	Good	Fail	
Assignment	Yes	21/21	5/5	0/4	26/30
	No	0/21	0/5	4/4	4/30
		21/30	5/30	4/30	

Likelihood Table		Grade			
		Excellent	Good	Fail	
Test	Pass	21/21	5/5	0/4	26/30
	Fail	0/21	0/5	4/4	4/30
		21/30	5/30	4/30	
Likelihood Table		Grade			
		Excellent	Good	Fail	
Presentation	Good	20/21	0/5	0/4	20/30
	Avg	1/21	5/5	0/4	6/30
	Poor	0/21	0/5	4/4	4/30
		21/30	5/30	4/30	

We computed all possible individual probabilities conditioned on the target attribute (Grade) in Table IV.

TABLE IV Possible Probabilities

P(Grade=Excellent)	21/30=0.7
P(Atten=Good∣ Grade=Ex)	0.9523
P(Atten=Avg∣ Grade=Ex)	0.0476
P(Atten=Poor∣ Grade=Ex)	0
P(Aptitu=Good∣ Grade=Ex)	0.3333
P(Aptitu=Avg∣ Grade=Ex)	0.4286
P(Aptitu=Poor∣ Grade=Ex)	0.2381
P(Assig=Yes∣ Grade=Ex)	1
P(Assig=No∣ Grade=Ex)	0
P(Test=Pass∣ Grade=Ex)	1
P(Test=Fail∣ Grade=Ex)	0
P(Present=Good∣ Grade=Ex)	0.9523
P(Present =Avg∣ Grade=Ex)	0.0476
P(Present =Poor∣ Grade=Ex)	0

P(Grade=Good)	5/30=0.1667
P(Atten=Good∣ Grade=G)	0
P(Atten=Avg∣ Grade=G)	0.8
P(Atten=Poor∣ Grade=G)	0.2
P(Aptitu=Good∣ Grade=G)	0.8
P(Aptitu=Avg∣ Grade=G)	0.2
P(Aptitu=Poor∣ Grade=G)	0
P(Assig=Yes∣ Grade=G)	1
P(Assig=No∣ Grade=G)	0
P(Test=Pass∣ Grade=G)	1
P(Test=Fail∣ Grade=G)	0
P(Present=Good∣ Grade=G)	0
P(Present =Avg∣ Grade=G)	1
P(Present =Poor∣ Grade=G)	0

P(Grade=Fail)	4/30=0.133
P(Atten=Good∣ Grade=Fail)	0
P(Atten=Avg∣ Grade=Fail)	0
P(Atten=Poor∣ Grade=Fail)	1
P(Aptitu=Good∣ Grade=Fail)	0.25
P(Aptitu=Avg∣ Grade=Fail)	0
P(Aptitu=Poor∣ Grade=Fail)	0.75
P(Assig=Yes∣ Grade=Fail)	0
P(Assig=No∣ Grade=Fail)	1
P(Test=Pass∣ Grade=Fail)	0
P(Test=Fail∣ Grade=Fail)	1
P(Present=Good∣ Grade=Fail)	0
P(Present =Avg∣ Grade=Fail)	0
P(Present =Poor∣ Grade=Fail)	1

b. Testing Data

Sr.no.	Roll no.	Attend-ance	Apti- tude	Assign-ment	Test	Presentation	Grade
1	IT1	Good	Avg	Yes	Pass	Good	?
2	IT27	Poor	Good	No	Fail	Poor	?

Likelihood of "Excellent" on that IT1

= P (Atten=Good| Grade=Ex) * P(Aptitu=Avg| Grade=Ex)*

P (Assig=Yes| Grade=Ex) * P(Test=Pass| Grade=Ex) *

P (Present=Good| Grade=Ex) * P(Grade=Excellent)

= 0.9523* 0.4286*1*1* 0.9523* 0.7 = 0.27208

Similarly Likelihood of "Good" on that IT1 =0

Similarly Likelihood of "Fail" on that IT1 =0

Similarly Likelihood of "Excellent" on that IT27 =0

Similarly Likelihood of "Good" on that IT27 =0

Likelihood of "Fail" on that IT27

= P (Atten=Poor| Grade=Fail) * P(Aptitu=Good| Grade=Fail)

* P (Assig=No| Grade=Fail) * P(Test=Fail| Grade=Fail)
* P(Present =Poor| Grade=Fail * P(Grade=Fail)

= 1 * 0.25 * 1 * 1* 1* 0.133 = 0.03325

Probability of IT1 being Grade = "Excellent" is 0.27208 Which is greater than another probability? Hence, IT1 will be Grade to Excellent.

Probability of IT27 being Grade =" Fail" is 0.03325which is greater than another probability. Hence, IT27 will be Grade to Fail.

V. CONCLUSION

A classification model has been proposed in this study for predicting student's grades particularly for IT under graduate students. In this paper, the classification task is used on student database to predict the students division on the basis of previous database. As there are many approaches that are used for data classification, Naive Bayesian classifiers method is used here. Information's like Attendance, aptitude, test, Presentation and Assignment marks were collected from the student's previous database, to predict the performance at the end of the semester. Bayes classifier algorithms are used in this model and the accuracy of prediction is compared to find the optimal one. Finally, the naïve Bayes algorithm is selected as the best algorithm for prediction based on performance detail.

REFERENCES

[1] Osmanbegovic E., Suljic M. "Data mining approach for predicting student performance" Economic Review-Journal of Economics and Business. Volume 10(1) (2012)

[2] Behrouz, M, Karshy, D, Korlemeyer G, Punch, W. "Predicting student performance: an application of data Mining methods with the educational web-based system" Lon-capa. 33rd ASEE/IEEE Frontiers in Education Conference. Boulder C.O. USA, (2003).

[3] Bekele, R., Menzel, W. "A bayesian approach to predict performance of a student (BAPPS): A Case with Ethiopian Students". Journal of Information Science (2013).

[4] Han and M. Kamber, "Data Mining: Concepts and Techniques, Morgan Kaufmann, 2000

[5] [5] Surjeet K, Yadav, Bharadwaj, B. Pal B." Data Minig Applications: A comparative Study for Predicting Student's performance." International journal of innovative technology & creative engineering. Volume 1(12). (2012).

[6] Nnamani, C. N, Dikko, H. G and Kinta, L. M. "Impact of students" financial strength on their academic performance: Kaduna Polytechnic experience". African Research Review 8(1), (2014).

[7] Ogunde A.O., Ajibade D.A. "A data Mining System for Predicting University Students F=Graduation Grade Using ID3 Decision Tree approach", Journal of Computer Science and Information Technology, Volume 2(1) (2014).

[8] Undavia, J. N., Dolia, P. M.; Shah, N. P. "Prediction of Graduate Students for Master Degree based on Their Past Performance using Decision Tree in Weka Environment". International Journal of Computer Applications; Volume 74 (21), (2013). https://www.wekatutorial.com/

[9] B. K. Baradwaj and S. Pal, "Mining educational data to analyze students' performance " *International Journal of Advanced Computer Science and Applications,* vol. 2, 2011.

[10] A. El-Halees, "Mining student data to analyze learning behavior: A case study," presented at the International Arab Conference of Information Technology (ACIT2008), Tunisia, 2008.

[11] George Dimitoglou, James A. Adams, and Carol M. Jim," Comparison of the C4.5 and a Naïve Bayes Classifier for the Prediction of Lung Cancer Survivability", Journal of Computing, Volume 4, Issue 8, 2012.

[12] Leung, K. Ming. "Naïve Bayesian classifier." Polytechnic University Department of Computer Science/Finance and Risk Engineering (2007).