A Comparative Analysis of Techniques for Predicting Academic Performance

Nguyen Thai Nghe[1], Paul Janecek[2], and Peter Haddawy[3]
Computer Science and Information Management Program
Asian Institute of Technology (AIT), Thailand 12120

Abstract - **This paper compares the accuracy of Decision Tree and Bayesian Network algorithms for predicting the academic performance of undergraduate and postgraduate students at two very different academic institutes: Can Tho University (CTU), a large national university in Viet Nam; and the Asian Institute of Technology (AIT), a small international postgraduate institute in Thailand that draws students from 86 different countries. Although the diversity of these two student populations is very different, the data-mining tools were able to achieve similar levels of accuracy for predicting student performance: 73/71% for {fail, fair, good, very good} and 94/93% for {fail, pass} at the CTU/AIT respectively. These predictions are most useful for identifying and assisting failing students at CTU (64% accurate), and for selecting Very Good students for scholarships at the AIT (82% accurate). In this analysis, the Decision Tree was consistently 3-12% more accurate than the Bayesian Network. The results of these case studies give insight into techniques for accurately predicting student performance, compare the accuracy of data mining algorithms, and demonstrate the maturity of open source tools.**

Index Terms - Bayesian Networks, Data Mining, Decision Trees, Prediction.

INTRODUCTION

Accurately predicting student performance is useful in many different contexts in universities. For example, identifying exceptional students for scholarships is an essential part of the admissions process in undergraduate and postgraduate institutions, and identifying weak students who are likely to fail is also important for allocating limited tutoring resources. This paper investigates the suitability of data mining tools for predicting academic performance using two case studies. In the first case study, Can Tho University (CTU) in Viet Nam, we use student records and GPA at the end of the 2nd year to predict performance in the 3rd year. In the second case study, the Asian Institute of Technology (AIT) in Thailand, we use admissions information, such as academic institute and GPA, to predict GPA at the end of the first year. This paper compares the accuracy of Decision Trees and Bayesian Network algorithms for predicting student performance in these two very different case studies.

This research has several important contributions. First, our results provide insight into the entire process of applying data mining tools to real-world data sets, including methods for refining data and improving prediction accuracy. Second, the results from these case studies show that the Decision Tree algorithm was significantly more accurate than the Bayesian Network algorithm for predicting student performance, based on the unmodified implementations provided by the Weka open source data-mining tool [1].

In the following section we describe the overall methodology of the research, from selection of a data-mining platform to modeling of the academic performance prediction problem. Next, we compare the results of the two prediction algorithms, followed by a comparison of our results with related work. Finally, we discuss the practical importance of this research, and our conclusions.

METHODOLOGY

This section describes the process we followed to collect and analyze the academic performance data. We discuss our selection of a data-mining tool, followed by the difficult task of preparing the data for analysis. We then present our model of the academic performance prediction problem, and how we tuned the parameters of the prediction algorithms to improve our initial results.

I. Selecting a data-mining tool

First, we conducted a detailed comparison of data mining tools to select a suitable platform for conducting our research. We began with a list of 30 data mining tools, which we filtered down to 10 with good support for visual analysis. We then applied the detailed methodology suggested by [2] to identify a number of computational, functional, usability, and support criteria necessary for this project. From the computational perspective, we required that the system operate on a wide range of platforms and be open source. This reduced the short list further to three: Weka [1], Orange [3] and Yale [4].

[1] Nguyen Thai Nghe, previously Master's Student, AIT, Thailand, and currently Lecturer, Can Tho University, Viet Nam, ntnghe@cit.ctu.edu.vn.
[2] Paul Janecek, Assistant Professor, AIT, Thailand, paul_janecek@ait.ac.th
[3] Peter Haddawy, Professor, AIT, Thailand, haddawy@ait.ac.th

Functionally, Weka and Yale support a wider range of algorithms than Orange, and have better data preparation tools. We eventually decided to use Weka based on its support for very large data sets.

II. Preparing the data

The next step was to gather, analyze and prepare the historical data from the academic records of the two institutes. For CTU, we were able to collect 20,492 complete records for students admitted from 1995 to 2002. For the AIT, we were able to collect 936 complete records for students admitted from 2003 to 2005. Figure 1 shows a distribution of the data that we would like to predict: the student's actual GPA at the end of the 3rd year of undergraduate at CTU, and at the end of the 1st year of postgraduate at AIT. The colors { ■, ■, ■, ■ } in the figure represent the classes {fail, fair, good, very good}.

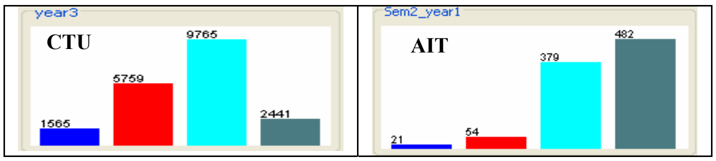


FIGURE 1
DISTRIBUTION OF ACTUAL GPA FOR CTU (LEFT) AND AIT (RIGHT)

In the data preparation phase we selected the relevant attributes from the available data, created meaningful groups within the attributes, and derived new attributes from our knowledge of the domain. Table I and II show summaries of the main attributes of the CTU and AIT data sets, and the "Information Gain" of each attribute for predicting academic performance. The information gain with respect to a set of examples is the expected reduction in entropy that results from splitting a set of examples using the values of that attribute. This measure is used in Decision Tree induction and is useful for identifying those attributes that have the greatest influence on classification.

For the CTU data set shown in Table I, the two attributes with the highest information gain are the Cumulative Grade Point Average for the 2nd year (*CGPA2*), and *English Skill*. The teaching and social environment at CTU is entirely in Vietnamese, but one possible explanation for the importance of English for predicting academic performance is the greater accessibility of information in English on the Internet.

For the AIT data set shown in Table II, the attribute with the highest information gain is Institute Rank, which is a derived attribute. One particularly challenging problem in preparing the AIT data was how to compare the Grade Point Averages (GPAs) of students from 329 institutes and over 40 different countries. There are two parts to this problem. First, the institutes have different grading scales (e.g., [0-10], [1-5], [0%-100%], [0-4]), which we convert into the 4.0 scale commonly used in North America. Second, and more challenging, each institute has a different grade distribution policy, which makes it difficult to equate the same GPA from different institutes. For example, the graduates with high GPAs from some institutes consistently graduate with lower GPAs from our institute.

TABLE I
CTU ATTRIBUTES, INFORMATION GAIN, AND RELATIVE CONTRIBUTION

#	Attribute	# Val	Values	Info Gain	Rel Gain
1	CGPA Year2	4	{Fail, Fair, Good, Very Good}	0.425	44.4%
2	English Skill	4	{A, B, C, N}	0.207	21.6%
3	Field of Study	18	{Accounting- Finance, … }	0.081	8.4%
4	Faculty	7	{Agriculture, …}	0.067	7.0%
5	Gender	2	{M, F}	0.064	6.7%
6	Entry Mark Range	4	{5.0-11.0, 11.5-14.0, 14.5-18.0, 18.5-30.0}	0.043	4.5%
7	Age Range	4	{15-17, 18, 19, 20-40}	0.020	2.1%
8	Policy Priority	2	{Yes, No}	0.012	1.2%
9	Area Priority	2	{Yes, No}	0.011	1.1%
10	Institute Rank	10	{1, 2, …, 10}	0.011	1.1%
11	Province	11	{AnGiang, BenTre, CaMau-BL, CanTho, DongThap, HauGiang, KienGiang, SocTrang, TienGiang, VinhLong, Others}	0.010	1.0%
12	Family Job	7	{Aquaculture, Business, Employee, Farmer, Gardener, Worker, Others}	0.007	0.7%
13	Ethnic	2	{KINH, OTHERS}	0.000	0.0%
14	Religion	2	{No, Yes}	0.000	0.0%

TABLE II
AIT ATTRIBUTES, INFORMATION GAIN, AND RELATIVE CONTRIBUTION

#	Attribute	# Val	Values	Info Gain	Rel Gain
1	Institute Rank	10	{1, 2, …, 10}	0.046	24.5%
2	Country	13	{Bangladesh, Cambodia, India, Indonesia, Laos, Myanmar, Nepal, Pakistan, PRChina, SriLanka, Thailand, Vietnam, Others}	0.035	18.9%
3	Entry GPA	4	{2.0-3.0, 3.0-3.3, 3.3-4.0}	0.019	10.2%
4	English Proficiency	4	{TOEFL, Certificate, Other, No}	0.017	9.3%
5	Donor	8	… scholarship providers …	0.016	8.5%
6	Current FOS	10	{ICT group, … , Others}	0.011	6.1%
7	Previous FOS	8	{IT group, …, Others}	0.008	4.3%
8	Current School	3	{Engineering, Resources and Development, Management}	0.008	4.3%
9	Fund Category	4	{Fellowship, Scholarship, Self-Support, Others}	0.007	3.8%
10	Marital Status	2	{Married, Single}	0.006	3.4%
11	Gross National Income	3	{Lower, Middle, Upper}	0.006	3.2%
12	Age Range	4	{20-24, 25-26, 27-30, 31-50}	0.005	2.9%
13	Gender	2	{M, F}	0.001	0.4%
14	TOEFL	3	{500-550, 551, 590, 591-677}		

To offset these differences between institutes, we derived a "rank" for each institute on a scale of 1 to 10. First, we assume that, on average, the grades that students receive upon graduation from our institute should be the same as the grades they received at their previous institute. We calculate the average difference of each institute from this assumption using the following equation:

$$Diff_{Institute} = AVG_{Institute}(GPA_{Graduation} - GPA_{Entry}) \qquad (1)$$

We then derive a "rank" for each institute by normalizing the resulting range of differences between institutes and scaling them from 1 to 10. The resulting distribution of *Institute Rank* for CTU and AIT is shown in Figure 2. The distribution on the left has already been rounded to the rank value, and the distribution on the right is still continuous. The colors correspond to the classes in Figure 1. For example, blue indicates those students who will have a failing grade, red indicates those with a fair grade, etc. It is important to note that this rank reflects differences in grading policies, not necessarily differences in education quality.

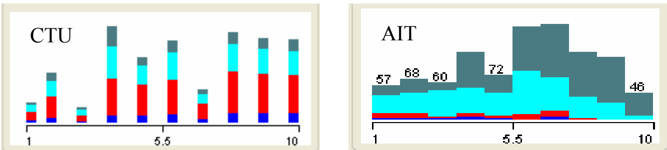


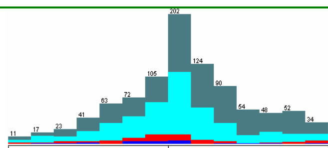
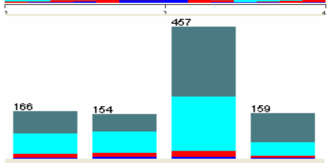
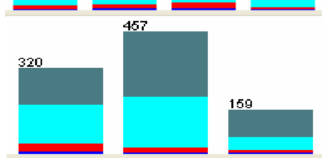
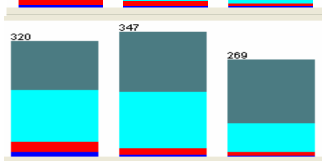
FIGURE 2
DISTRIBUTION OF INSTITUTION RANKS FOR CTU (LEFT) AND AIT (RIGHT)

As shown in Table II, the derived *Institute Rank* attribute had the highest information gain for the AIT data set, and provided 2.5 times more information gain than the *Entry GPA* for predicting student performance. *Institute Rank* was much less important for the CTU data set than for the AIT data set, possibly because CTU uses a standardized test, *Entry Mark*, rather than *Entry GPA*.

III. Modeling the academic performance prediction problem

The next step in the academic performance prediction problem was to construct and evaluate models with the Decision Tree and Bayesian Network algorithms available in the Weka data-mining tool. Using these models, we "tuned" the input attributes of the data sets by subdividing the range of values into new classes and evaluating the changes in the accuracy of prediction. In some cases, this led to significant improvements in accuracy. For example, we experimented with several divisions of the continuous attribute *Entry GPA* as shown in Table III.

TABLE III
TUNING THE VALUES OF THE "ENTRY GPA" ATTRIBUTE FOR THE AIT

Solution	Data Distribution	%Correct - 4 classes
Continuous [2.0…. 4.0]		58.95 %
{ Fail, Fair, Good, VeryGood}		57.08 %
{ Warning, Good, VeryGood}		58.22 %
{ 2.0-3.0, 3.0-3.3, 3.3-4.0}		60.27 %

The range of the continuous distribution shown in the first row of Table III is 2.0 to 4.0, with the largest number of students at 3.0. The 4-category solution divides the range into {Fail, Fair, Good, Very Good} groups (i.e., {C, C+, B, B+/A} or {2.0-2.5, 2.5-3.0, 3.0-3.5, 3.5-4.0}), and results in a slightly lower accuracy for prediction. The third solution grouped the 2.0-3.0 students into a single group, with slightly better results, but still lower than the continuous distribution. The fourth solution divides the students at 3.3 rather than 3.5, resulting in three similarly sized groups and a 1.3% higher accuracy for prediction than with the continuous distribution.

We analyzed each attribute to determine whether grouping the data would improve the accuracy, and found slight (<1%) improvements in the CTU data set for *Religion, Entry Mark, Area Priority, Policy Priority.* We tried several divisions of *English Proficiency* and *TOEFL* score, but they did not lead to improvements over the basic classification of {TOEFL, Certificate, Other – Yes, No}. The final results of the classifications are the values shown in Tables I and II.

IV. Tuning the Parameters of the Algorithms

We compared three of the prediction algorithms provided by Weka: the J48 Decision Tree, the M5P Model Tree, and the BayesNet Bayesian Network. In this section we describe the configuration of these algorithms for our experiments.

The J48 Decision Tree algorithm in Weka is available in the Java class "weka.classifiers.trees.J48". This class generates a pruned or un-pruned C4.5 Decision Tree, and has the following three main parameters (see Table IV):

- `confidenceFactor`: The confidence factor used for pruning (smaller values incur more pruning)
- `minNumObj`: The minimum number of instances per leaf.
- `Unpruned`: Whether pruning is performed or not.

TABLE IV
PARAMETER VALUES FOR THE DECISION TREE

Predicted Values: GPA Classes	Parameters	CTU	AIT
4: {Fail, Fair, Good, Very Good}	`confidenceFactor`	0.25	0.25
	`minNumObj`	2	3
	`Unpruned`	False	False
3: {Fail, Good, Very Good}	`confidenceFactor`	0.25	0.25
	`minNumObj`	2	3
	`Unpruned`	False	False
2: {Fail, Pass}	`confidenceFactor`	0.25	0.25
	`minNumObj`	2	7
	`Unpruned`	False	True

The M5P Model Tree algorithm in WEKA is available in the Java class "weka.classifiers.trees.M5P". The two main parameters are described below (see Table V):

- `buildRegressionTree`: If True then the algorithm builds a regression tree rather than a model tree.
- `minNumInstances`: The minimum number of instances to allow at a leaf node.

TABLE V
PARAMETER VALUES FOR THE MODEL TREE

Predicted Values: GPA Classes	Parameters	CTU	AIT
Numeric	`buildRegressionTree`	False	
	`minNumInstances`	4	

The Bayesian Network algorithm in Weka is available in the Java class "weka.classifiers.bayes.BayesNet", and has the following two main parameters (see Table VI):

- `Estimator:` The algorithm used for finding the conditional probability tables. The SimpleEstimator estimates the probabilities directly from data.
- `SearchAlgorithm:` The method used for searching the network structures.

TABLE VI
PARAMETER VALUES FOR THE BAYESIAN NETWORK

Predicted Values: GPA Classes	Parameters	CTU	AIT
4, 3, and 2	`Estimator`	`SimpleEstimator`	
	`SearchAlgorithm`	`HillClimbing`	

RESULTS AND ANALYSIS

A summary of the results of the CTU and AIT predictions using the Decision Tree (DT) and Bayesian Network (BN) algorithms is shown in Table VII. Tables VIII through XII show the detailed results from the Decision Tree predictions for each 4-, 3-, and 2-classes. The CTU prediction is for the CGPA of undergraduate students at the end of their 3rd year given data from the end of their 2nd year, and the AIT prediction is for Masters students at the end of their 1st year given their admissions information. The accuracy of the prediction is evaluated using Cross-Validation with 10 folds. The results for CTU and AIT in Tables VII through XII are shown with both the original data and the resampled data. The purpose and technique for resampling is explained in the next section.

The summary results in Table VII show that the predictions for the CTU data set are noticeably more accurate than for the AIT data set, which is expected given the much larger number of records for CTU. The results also show the Decision Tree algorithm consistently outperformed the Bayesian Network algorithm (e.g., up to 12% in the 4-classes case).

TABLE VII
COMPARISON OF GPA PREDICTION RESULTS FOR CTU AND AIT

Predicted GPA Classes	Algo	CTU (20,492 records) %Accuracy Original Data	CTU (20,492 records) %Accuracy Resampled Data	AIT (936 records) %Accuracy Original Data	AIT (936 records) %Accuracy Resampled Data
4 classes:	DT	66.69%	**72.95%**	63.25%	**70.62%**
{Fail, Fair, Good, Very Good}	BN	61.32%	60.80%	57.48%	61.54%
3 classes:	DT	84.18%	**86.47%**	67.74%	**74.36%**
{Fail, Good, Very Good}	BN	78.57%	78.73%	63.89%	66.13%
2 classes:	DT	92.86%	**94.03%**	91.98%	**92.74%**
{Fail, Pass}	BN	89.75%	90.27%	90.91%	88.57%

TABLE VIII
DECISION TREE CONFUSION MATRIX, CTU CASE STUDY, 4 CLASSES

Actual Class	Predicted Class (CTU) Original Data: Fail	Fair	Good	Very Good	Re-Sampled Data: Fail	Fair	Good	Very Good
Fail	534	890	134	7	849	543	169	4
Fair	360	3499	1888	12	336	3757	1609	23
Good	30	1519	7701	515	83	1093	8214	407
Very Good	1	15	1135	1290	6	32	977	1428
% Hit	34 %	61 %	79 %	53 %	**55 %**	**66 %**	**84 %**	**58 %**

TABLE IX
DECISION TREE CONFUSION MATRIX, AIT CASE STUDY, 4 CLASSES

Actual Class	Predicted Class (AIT) Original Data: Warn	Fair	Good	Very Good	Re-Sampled Data: Fail	Fair	Good	Very Good
Fail	0	1	14	6	4	1	12	1
Fair	0	1	43	10	1	18	17	14
Good	1	1	247	130	7	10	222	119
Very Good	1	9	129	344	3	11	79	417
% Hit	0%	2%	65%	71%	**22%**	**36%**	62%	**82%**

TABLE X
DECISION TREE CONFUSION MATRIX, CTU CASE STUDY, 3 CLASSES

Actual Class	Predicted Class (CTU) Original Data: Fail	Good	Very Good	Re-Sampled Data: Fail	Good	Very Good
Fail	468	1090	7	579	984	2
Good	312	14707	505	177	14985	360
Very Good	0	1174	1267	3	1117	1323
% Hit	*30%*	*95%*	*52%*	***37%***	***97%***	***54%***

TABLE XI
DECISION TREE CONFUSION MATRIX, AIT CASE STUDY, 3 CLASSES

Actual Class	Predicted Class (AIT) Original Data: Fail	Good	Very Good	Re-Sampled Data: Fail	Good	Very Good
Fail	0	15	6	1	13	4
Good	1	302	130	2	298	108
Very Good	1	150	332	3	110	397
% Hit	*0%*	*70%*	*69%*	***5%***	***73%***	***78%***

TABLE XII
DECISION TREE CONFUSION MATRIX, CTU AND AIT CASES, 2 CLASSES

Actual Class	Predicted Class CTU Data Original: Fail	Pass	Re-Sampled: Fail	Pass	AIT Data Original: Fail	Pass	Re-Sampled: Fail	Pass
Fail	471	1094	997	568	2	73	32	36
Pass	300	17665	526	17439	13	848	32	836
% Hit	30 %	98 %	**64 %**	97 %	2 %	98 %	**47 %**	96 %

I. Re-sampling

Upon closer inspection of the results using the confusion matrices in Tables VIII through XII, we found that there were large imbalances in the distributions of the output classes, and that the accuracy in the smaller classes was much lower than the accuracy in the larger classes. For example, the number of students predicted to fail for the CTU data set (see Table VIII, "Original Data") is nearly 15 times smaller than the number predicted to receive "Good", and the accuracies for these two classes are 34% and 79%, respectively. The prediction accuracies for the smaller (i.e. minority) classes are consistently lower for both data sets and for all classes.

To compensate for this problem, we used the *resample* function in Weka, which oversamples the minority class and undersamples the majority class to create a more balanced distribution for training the algorithms. Table VII shows the accuracy of the predictions when training with both the original and re-sampled data sets. The predictions using the re-sampled data set are significantly more accurate.

II. Detailed Analysis of Distributions with Confusion Matrices

The confusion matrices for prediction into 4-, 3-, and 2-classes (Tables VIII through XII) show the changes in the distributions of predicted and actual values for both the original and re-sampled data. The accuracy of the prediction using the re-sampled data sets significantly improves where the original data has a much smaller sample size (for an overview of the actual values, see the histograms in Figure 1). For example, re-sampling improved the accuracy of the prediction of which students would fail from 34% to 55% in the CTU case (Table VIII), and 0% to 22% in the AIT case (Table IX) using the Decision Tree algorithm. Similarly, the accuracy of the 2-class prediction for failing students increased from 30% to 64% for the CTU case and 2% to 47% for the AIT case (Table XII).

Table VII shows that the highest accuracies are achieved in the 2-classes case. However, this is largely a product of the small percentage of students who fall into the *Fail* class compared to the *Pass* class. Careful inspection of the distributions in the confusion matrices reveals that the highest accuracies are always achieved for the largest classes, which is the *Good* students in CTU (undergraduate) and the *Very Good* students at the AIT (postgraduate).

The 3-classes case is more accurate overall than the 4-classes case (as shown in Table VII), but less accurate in the prediction of the *Fail* and *Very Good* classes that are most useful for the academic context (as shown in Tables VIII through XI). The 2-classes case was the most accurate for predicting *Fail* for both data sets (CTU, 64% and AIT, 47%), and the 4-classes case was the most accurate for predicting *Very Good* for both data sets (CTU, 58% and AIT, 82%). The accuracy of these results is certainly sufficient for guiding decision makers in allocating their limited resources.

RELATED WORK

Data mining has been used for many different purposes in academics. Given the large amounts of data collected in academic institutions, [6] proposes a model with different types of education-related questions and the data-mining techniques appropriate for them. For example, predicting student performance, clustering similar students, and associating types of students with appropriate courses. An example of a specific case study of clustering students with similar characteristics (such as "self starters" and "high interaction") is given in [5].

We compare our research with two examples of research that also attempt to predict student performance [7,8] in Table XIII. We can see that the size of our data set was much larger than these previous studies. Our overall results in the 3-classes prediction were slightly more accurate than [8] with the AIT case, and over 14% more accurate with the CTU case. In the 2-classes prediction, [8] performed 3-4% better. However, given that the data sets are completely different in size and number of attributes, this comparison is primarily to give an appreciation of the different approaches to the problem of predicting student performance and the typical accuracy.

TABLE XIII
COMPARISON WITH RELATED WORK

Criteria	This Research		Previous Research	
	CTU	AIT	[7]	[8]
Prediction Problem	Student GPA		Mathematics / English performance	Final grade of Physics course
Student type	Undergraduate	Graduate	High School	Undergraduate
Data set	20,492	936	514	261
Number of attributes	15	15	8	-
Predict attribute values	• 4-classes (Fail, Fair, Good, Very Good), • 3-classes (Fail, Good, Very Good), • 2-classes (Fail, Pass), Numeric GPA		• 3-classes (Below, Satisfactory, Above)	• 9-classes (0.0,..,4.0) • 3-classes (High, Middle, Low) • 2-classes (Fail, Pass)
Techniques	Decision Tree/Model Tree		Bayesian Network	Genetic Algorithm
Accuracy percentage (4-classes)	72.95 %	70.62%		62.88 % (9-classes)
(3-classes)	86.47 %	74.36 %	64 %	72.52 %
(2-classes)	94.03 %	92.74 %		96.93%
System type	Web-based		Application	Application
Platform	Weka		BNJ, Weka	MATLAB

DISCUSSION AND CONCLUSIONS

Predictions of student performance can be useful in many contexts. For admissions, it is important to be able to identify excellent students for allocating scholarships and fellowships, as well as those students who are unlikely to graduate. This task is extremely difficult with international students, who come from institutions with diverse grading systems and have backgrounds that faculty and staff are often unfamiliar with. The overall prediction accuracy from our analysis was 86%

(CTU) and 74% (AIT) for the 3-class prediction. In the case of admissions decisions for the AIT, the accuracy for predicting excellent (i.e., B+/A) students was as high as 82%, and the accuracy for identifying students who would likely fail was 47%. Therefore, while the system is reliable for identifying excellent students for the AIT, we will need to continue to work on the classification problem for identifying students who are most likely to fail.

We have implemented and deployed this system as a web-based application for the faculty at the AIT. Although Weka was not designed to run over the web, as an Open Source application, we were able to extract the necessary classes and wrap them in a web-based interface.

ACKNOWLEDGMENT

This research was conducted as part of the Master's thesis of the first author, and was funded by a scholarship from the Ministry of Education, Viet Nam.

REFERENCES

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

[2] Collier, K., Carey, B., Sautter, D., and Marjaniemi, C., "A methodology for evaluating and selecting data mining software," in *Proceedings of the 32nd Annual Hawaii International Conference on System Sciences*, IEEE, 1999.

[3] Orange, University of Ljubljana, Slovenia, http://www.ailab.si/orange

[4] Yale, Rapid-I, http://rapid-i.com/

[5] Luan, J., Zhao, C.-M., and Hayek, J., "Exploring a new frontier in higher education research: A case study analysis of using data mining techniques to create NSSE institutional typology," *Paper presented at the California Association for Institutional Research*, Anaheim, California, November 17-19, 2004.

[6] Delavari N. & Beikzadeh M. R & Shirazi M. R. A., "A New Model for Using Data Mining in Higher Educational System", in *Proceedings of 5th International Conference on Information Technology Based Higher Education and Training (ITHET),* Istanbul, Turkey, May 31 to June 2, 2004.

[7] Bekele, R. and Menzel, W., "A Bayesian approach to predict performance of a student (BAPPS): A case with Ethiopian students," in *Proceedings of the International Conference on Artificial Intelligence and Applications (AIA-2005)*, Vienna, Austria, 2005.

[8] Minaei-Bidgoli, B., Kashy, D. A., Kortemeyer, G., and Punch, W. F, "Predicting student performance: an application of data mining methods with an educational web-based system," in *Proceedings of 33rd Annual Conference on Frontiers in Education (FIE 2003)*, volume 1, 2003, pages 13–18.