The study of Factors and Model Creation for Predicting Academic Achievement Using Support Vector Machine

Thidarat Pinthong
Department of Business Information Management
Phetchaburi Rajabhat University
Thailand
Phetchaburi , Thailand
thidarat.pin@mail.pbru.ac.th

Worawut Yimyam
Department of Business Information Management
Phetchaburi Rajabhat University
Thailand
Phetchaburi , Thailand
worawut_yimyam@hotmail.com

Mahasak Ketcham
Department of Information Technology Management
King Mongkut's University of Technology North Bangkok, Thailand
Bangkok, Thailand
mahasak.k@it.kmutnb.ac.th

Abstract—This research proposed the study of factors and model creation for predict-ing academic achievement and managing the teaching to suit student groups. The clustering Technique and Support Vector Machine technique was applied to this research. The results of the experiment show that 1) the major factors affecting the achievement include 6 factors the pre-study results which are residence, sex, family background, learning outcomes before ad-mission and selected majors 2) the classification by comparing the classifica-tion models, it was found that SVM forecasting had the best performance (Accuracy = 92.93, Recall = 0.929 and Precision = 0.929)

Keywords—academic achievement, clustering Teaching, Support Vector Machine

I. INTRODUCTION

Nowadays, education is essential as it provides knowledge and develops the country's citizen. As a result, quality of higher education institutions have been developed in many aspects including advanced curriculum development to serve labor market, teacher quality development and learner quality development [1] etc. This is related to learner academic achievement. This achievement is success of learners gained from learning. Therefore, it is essential to study factors influencing academic achievement. This study analyzed learners' personal information in higher education institutions' information systems to seek factors influencing learner academic achievement in order to cluster and classify students to organize suitable teaching and learning.

This study aimed to study factors and to create model for predicting academic achievement. The study was divided into 2 parts: clustering by DBScan to find relationship among groups that identifies factors influencing academic achievement and; classification by Support Vector Machine. Then, Decision Trees, Baysian Naïvebayes, Artificial Neural Network and Fuzzy were applied to make a comparison. Learners' information was classified into 2 groups including learners who achieved academic achievement and learners who did not achieve academic achievement.

II. PRELIMINARY

A. Learning Achievement

Development of education quality and standard is an important part of educational reform. Nowadays, The National Education Plan Revised Version (B.E. 2552 - 2559) [2] defines educational quality and learning development in all levels of education including development of teachers' knowledge and curriculum development which maximize learners' efficiency. The important learning factors for learners are teachers who influence knowledge, concepts and methods.

III. PROPOSED METHOD

This research aims to study the factors affecting student achievement and create model for forecasting using by the Support Vector Machine method and comparison performance as shown in Figure 1 as follows.

A. Data of Research

This study uses the dataset of personal of student, data results of Senior High School, Achievement information of bachelor degrees for faculty of management science in academic year 2551 to 2555.It have data amount 841 record. The details are shown in Table 1.

TABLE I.   DETAIL OF DATA STUDENT

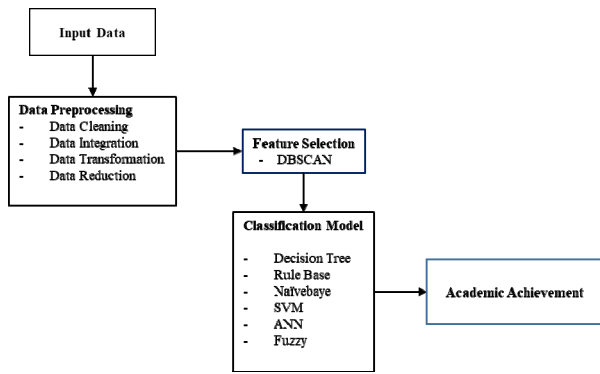| No	Name	Detail	Type
1	Sex	Sex	Input
2	Home	The Shelter	Input
3	Sta_parents	Family status	Input
4	Disease	Congenital disease	Input
5	Deformity	Disability	Input
6	Old_edu	Qualifications applied	Input
7	Old_grade	data results of Senior High School	Input
8	Major	Major	Input
9	Graduation	Educational Achievement	Output

Figure 1. Framework for studying factors and creating model for predicting academic achievement

B. Data Preprocessing

It prepares data before producing data mining. This step includes, Data Cleaning, Data Integration, Data Transformation and Data Reduction to have ac-curate and appropriate data for clustering and classification.

C. Feature Selection

It selects data influencing academic achievement using personal information, learning outcomes before admission and learning achievement to seek factor relationships in different dimensions. Data is clustered by DBScan Algorithm to find factor relationship influencing learners' achievement.

D. Comparison Classification Model

The researcher designed an experiment to test efficiency by considering Accuracy, Precision and Recall as illustrated in Equation (1) (2) and (3) as follows;

$$Precision(p) = \frac{TP}{TP + FP} \qquad (1)$$

$$Recall(r) = \frac{TP}{TP + FN} \qquad (2)$$

$$Accuracy(A) = \frac{TP + TN}{TP + TN + FP + FN} \qquad (3)$$

When TP = True Positive, FP = False Positive, FT = True Negative และ FN = False Negative

IV. EXPERIMENTAL RESULTS

A. The Results of Factor Relationship

To identify relationships, clustering was applied with personal information, learning outcomes and learning achievement results using DBScan. The Epsilon and minPoints were defined to suit clustering. The appropriate values were: Epsilon = 0.8 and minPoints = 4. There were 55 clusters and could be grouped into 2 main clusters including achieving cluster and not-achieving cluster. In not-achieving cluster, factors could be identified from 6 sub-clusters in-

cluding learning outcomes before admission, residence, sex, family background, qualification before admission and selected major. This could be illustrated as in Table 1.

TABLE II.   FACTORS INFLUENCING ACADEMIC ACHIEVEMENT

Variable	Values	Probability
Learning outcomes before admission	Less than 2.97	0.8417
Residence	Own house	0.7907
Sex	Female	0.7860
Family background	Parents living together	0.6977
Qualification before admission	Grade 12	0.6744
Selected major	Business Computer	0.3814

B. The Results of General Model Efficiency of General Model

The results of creating a model to classify academic achievement through the 6 models give the efficiency value consisting of accuracy, precision and re-call as illustrated in Table 2.

TABLE III.   A COMPARISON OF GENERAL MODEL EFFICIENCY

Model	10-fold cross validation		
	Accuracy	Recall	Precision
DT	83.898	0.839	0.839
RB	79.378	0.794	0.818
BN	78.531	0.785	0.793
SVM	92.938	0.929	0.929
ANN	88.701	0.891	0.871
Fuzzy	89.266	0.897	0.877

V. CONCLUSION

This study aimed to create an efficient model to predict learner academic achievement by using personal information, learning outcomes before admission and learners' learning achievement. It was found that the significant factors influencing learning achievement included 6 factors consisting of learning outcomes before admission, residence, sex, family background, qualification before admission and selected major. The most efficient model for prediction was SVM (Accuracy = 92.93, Recall = 0.929 and Precision = 0.929).

There shall be more studies on other related factors expected to influence academic achievement to maximize the efficiency or apply other techniques.

REFERENCES

[1]   Smith, T.F., Waterman, M.S.: Identification of Common Molecular Subsequences. J. Mol. Biol. 147, 195–197 (1981)

[2] May, P., Ehrlich, H.C., Steinke, T.: ZIB Structure Prediction Pipeline: Composing a Com-plex Biological Workflow through Web Services. In: Nagel, W.E., Walter, W.V., Lehner, W. (eds.) Euro-Par 2006. LNCS, vol. 4128, pp. 1148–1158. Springer, Heidelberg (2006)

[3] Foster, I., Kesselman, C.: The Grid: Blueprint for a New Computing Infrastructure. Mor-gan Kaufmann, San Francisco (1999)

[4] Czajkowski, K., Fitzgerald, S., Foster, I., Kesselman, C.: Grid Information Services for Distributed Resource Sharing. In: 10th IEEE International Symposium on High Perfor-mance Distributed Computing, pp. 181–184. IEEE Press, New York (2001)

[5] Foster, I., Kesselman, C., Nick, J., Tuecke, S.: The Physiology of the Grid: an Open Grid Services Architecture for Distributed Systems Integration. Technical report, Global Grid Forum (2002)

[6] National Center for Biotechnology Information, http://www.ncbi.nlm.nih.gov