AN INTELLIGENT RECOMMENDATION SYSTEM FOR TEACHING

Michael Watkins[1*], Kanokwan Watkins[2] and Nutsana Napayap[3]

[1] Faculty of Arts, Murdoch University, Australia
[2] Dean, Didyasarin International College, Hatyai University
[3] Head, Department of Digital Media Design, Didyasarin International College, Hatyai Univeristy
[*] **Corresponding author, E-mail:** kanokwan@hu.ac.th

Abstract

Within Thailand, many universities face the problems of student dropouts or failures before graduations. In order to improve and support the academic management processes, some universities are developing innovative information systems and services with an aim to enhance efficiency and retain the students to graduations. Moreover, the information technological support can also improve student relationship with universities. One of the key initiatives is the development of Student Relationship Management Systems (SRM) and among their functions, is the provision of recommendation and advice for students. Intelligent Recommendation systems allow personalization for counseling. The objective of this research is to develop and apply intelligent techniques and methodologies to a recommendation system for recommending teaching styles to match with learning styles of students. This includes recommending activities and games for teachers in order to give lecture in the classroom. The proposed techniques will be implemented and evaluated based on classification models. The best model will be chosen to apply in the intelligent recommendation system for teaching, and a drop-out recommendation module is employed to the system.

Keywords: intelligent recommendation system, SRM, data mining.

Introduction

The Office of Higher Education Institutes in Thailand [1] has one of the key objectives of is to focus on improving student completion rates, along with enhancing graduates' skills in their respective programs of study In addition, to foster students' creativity and innovation, it is essential for the schools to be accountable for providing active curriculums as well as technological supports to enrich their learning outcomes and improve skills. In term of providing technologies to support teachers, supervisors and students, it is one of the strategies for school management. In order to focus on the improvement of students' completion rates from schools, there are various reasons why a student may choose to drop out. One of these issues is depression[2]. This can be from family problems or other reasons. However, it can occur that students leave the schools early.

With the advancement of information technologies, teaching styles can be enhanced through innovative technologies in the classroom, such as interactive classrooms that, in turn, inspire students and create an enthusiastic and positive learning environment. This inspiring learning atmosphere is likely to support students in terms of expanding their own innovation and creativity. Understanding student needs will enhance their learning experience, increase their chance of success and reduce resource wastage that is due to dropouts.

With a limited supply of resources and increasing competition for students entering to a course of study in university, some students fail to enter to the course of study, and some students could enter to the expected course. However, they are unable to cope with the study, which is common problem among tertiary students. Choosing the course of study for high school students is important. Provision of counseling services in schools is an factor contributing to students' academic success [3].

Likewise, developing a system to assist students and lecturers in the classrooms is also focused. This research study aims to investigate and develop an intelligent system to provide academic recommendations for students. Also, it will provide recommendations for teachers, counselors and student's supervisors in various aspects in terms of teaching and giving guidance for students. This will enhance students' learning experience, increase their chances of success, and reduce resources expenditure for students in the Southern Thailand.

Expected Benefits

7.1. The research will gather information from university students and lecturers in terms of types of learning and teaching styles including activities and games. This information can be used to support teaching.

7.2. The research outcome will provide an intelligent system for teaching for university students in Southern Thailand.

7.3. The research outcome will be able to apply for teaching in various courses.

Scope of Research

In this research, the respondents are students from 5 universities in Southern Thailand. The variables in this research are included independent and dependent variables. Independent variables are demographic information such as gender, age, domicile, education, incomes, types of learning behaviors, types of teaching styles, favorite games and activities in each types of classroom. On the other hand, dependent variables are matching type of learning behaviors and teaching style and Recommending of games and activities for teaching. In addition, Drop-out recommendation model is employed to use in the recommendation system.

Intelligent Recommendation System

Prior studies have investigated and developed the recommendation systems [4-6]. Kongsakun[5] proposed an intelligent recommendation system in order to improve student performances; two of the modules are program and activity recommendation. The activity recommendation module provided the ranked activities in university for students to enhance their skills and performances. These two modules used student historic data to correlate with student background information in order to provide the structural recommendation for students [5]. Hence, a system that recommends more appropriate program placement, leading to higher level of success, could be considered a high quality supporting service, thereby, increasing student retention and able to complete their study.

Furthermore, there are various types of learning objects (LOs) such as digital repository, eBook etc. available online for students, however; there is a problem with these learning objects' ability in searching or retrieving appropriate LOs for individual learners[7]. Having learning object recommendation is able to characterize the learning objects and match to student's personal styles and backgrounds is one of the educational data mining [8]. One of other learning style integration is an automatic learner modeling approach. This classification was used to classify the learners based on their interests[9]. With this in mind, learning object recommendation model will highlight and assist students in terms of searching appropriate recommendation. Prior studies have addressed issues faced by school teachers for classroom teaching. Teaching styles of teachers may not suit with student learning behaviors, therefore; students are lack of understanding and motivation to study. One of the Educational Data mining researchers views is predicting students' future learning behavior by creating student models[8]. In term of classroom recommendation, this prediction model incorporate student information in order to match student's information with teacher teaching styles[8]. As said before, one of the strategies in education is technological support. In classroom, the computational structural recommendations can be applied in various formats. The intelligent recommendation system will be developed in order to assist students, teachers, supervisors and counselors. This forms the motivation of this research.

Intelligent Techniques for the proposed recommendation system

With the context of recommendation systems, some studies focused on the improvement of recommendation systems [6, 10, 11]. In this study, various data mining techniques have been used such as ANN, SVM, DT, Clustering, AR etc. Some examples are explained below.

Decision Tree (DT), one of the data mining techniques, resembles an inverted tree structures. DT has been used to develop web application based on personalized requirement and enhanced efficiency of learning environment[12, 13]. Another popular technique is Artificial Neural Networks. A report by Kala et al.[14] has illustrated that ANNs and machine learning have been used in large number of research studies analyzing large datasets.[14, 15]. Next, Support Vector Machine (SVM), a well known classification techniques and supervised learning method, has been used to find the hidden relation models in various ways. For example, Barbella et.al [16] employed SVM to classify data for a recommendation system, whereas Martens et. al[17] extracted rules from SVM. Furthermore, various data mining techniques will be used in this study including Aggregation Techniques in order to find the highest accuracy for the intelligent recommendation system.

In order to actualize the proposed systems, various techniques and technologies will be included in the study. This research adopts statistics and data mining for analyzing the data. The main analysis tools used are R and Rapidminer. Moreover, PHP and MySQL will be used to develop the web-based applications.

Conceptual Framework

The proposed system framework is presented in Figure 1. The framework composes of three consecutive layers, Frontend, Backend, and Data and knowledge bases. Frontend represents the system’s user interfaces, which allow users to interact with the system. Backend represents the system’s core engines, which undertake the system functions. The bottom layer stores the databases and knowledge bases for the system. The frontend and the backend, in turn, the model consists of modules according to proposed system functions.

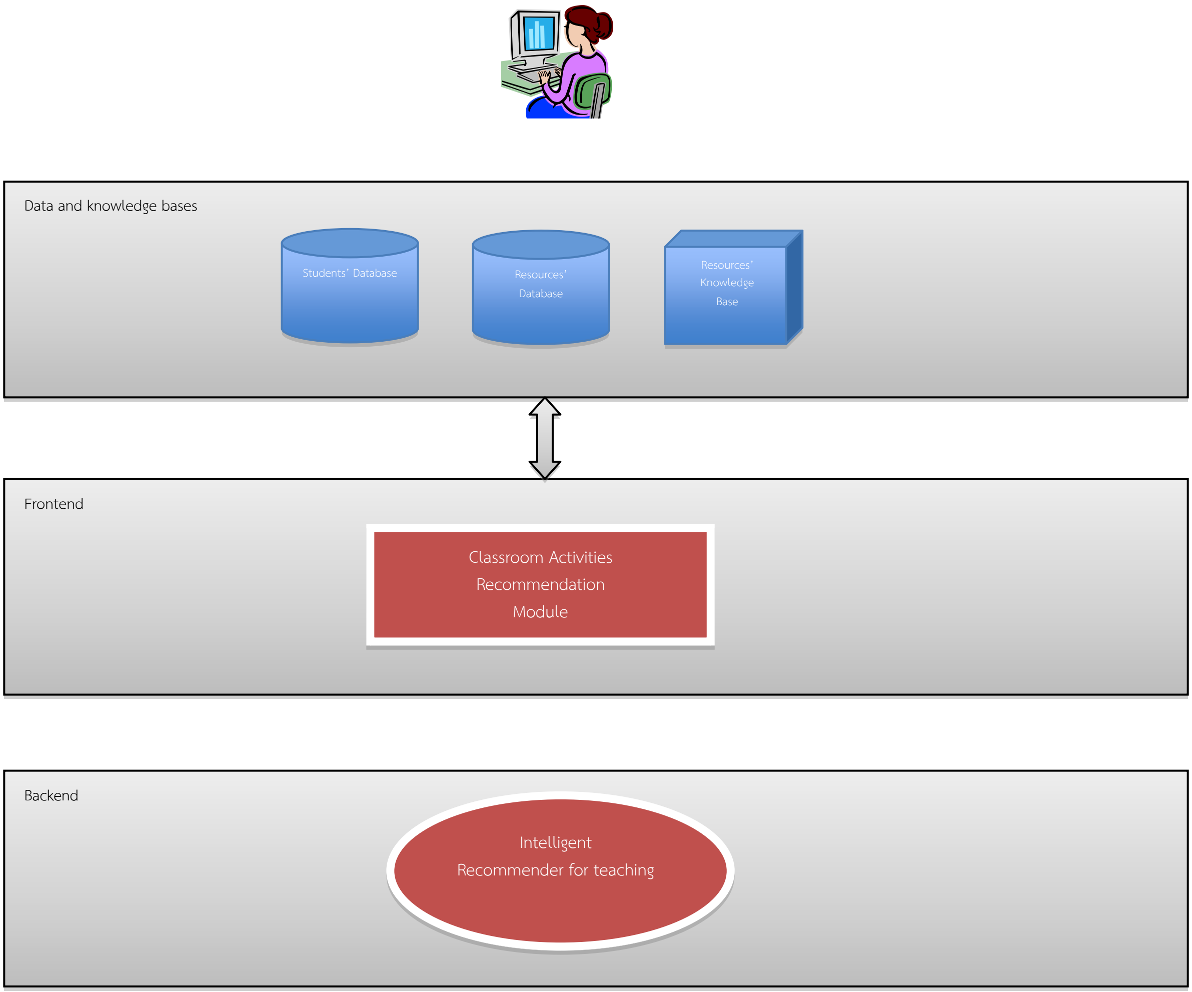

Figure 1 The proposed system model

In order to implement the system model in Figure 1, the essence process is how to create the databases and knowledge bases used by the systems. This part is very important because it enable the system to perform intelligently.

The data mining process, CRIPS_DM, is normally used to create those knowledge bases (and databases) for recommendation systems [18]. CRISP-DM conceptual model includes business understanding, data understanding, data preparation, modeling, evaluation, and deployment. Based on CRISP-DM conceptual model, the databases and knowledge bases used in the proposed system can be established as shown in Figure 2.

Data Analysis

As mentioned that databases and knowledge bases are the essence part of the proposed system, the substantial process to establish them, however, have not been described. Figure 2 describe the process to establish those database and knowledge bases. The first tier of the process is the data collection. Several ways include handing out questionnaire.

Once the data have been collected, they will be prepared into appropriately computational formats such as databases, text files, or spread sheets. These formats are suitable for R statistics and Rapid miner applications to perform preprocessing and preliminary analysis.

In the third tier, preprocessing and preliminary analysis, the data will be cleaned in order to increase usability and reliability. For examples, missing data will be removed or imputed, select necessary attributes, perform the transformation (i.e. normalization). Moreover, outliner detection and statistics test will be performed to remove noise from the data.

When the data are ready, they will be analyzed or modeled according the needs of the modules. The data will be used to perform correlation analysis, association analysis, and create classification models. For future activities recommendation module, the data are used to do clustering analysis, and create prediction or classification models.

Following step mentioned above, the databases and knowledge bases are ready for intelligent recommendation systems. The next step is to develop web-based applications for those models

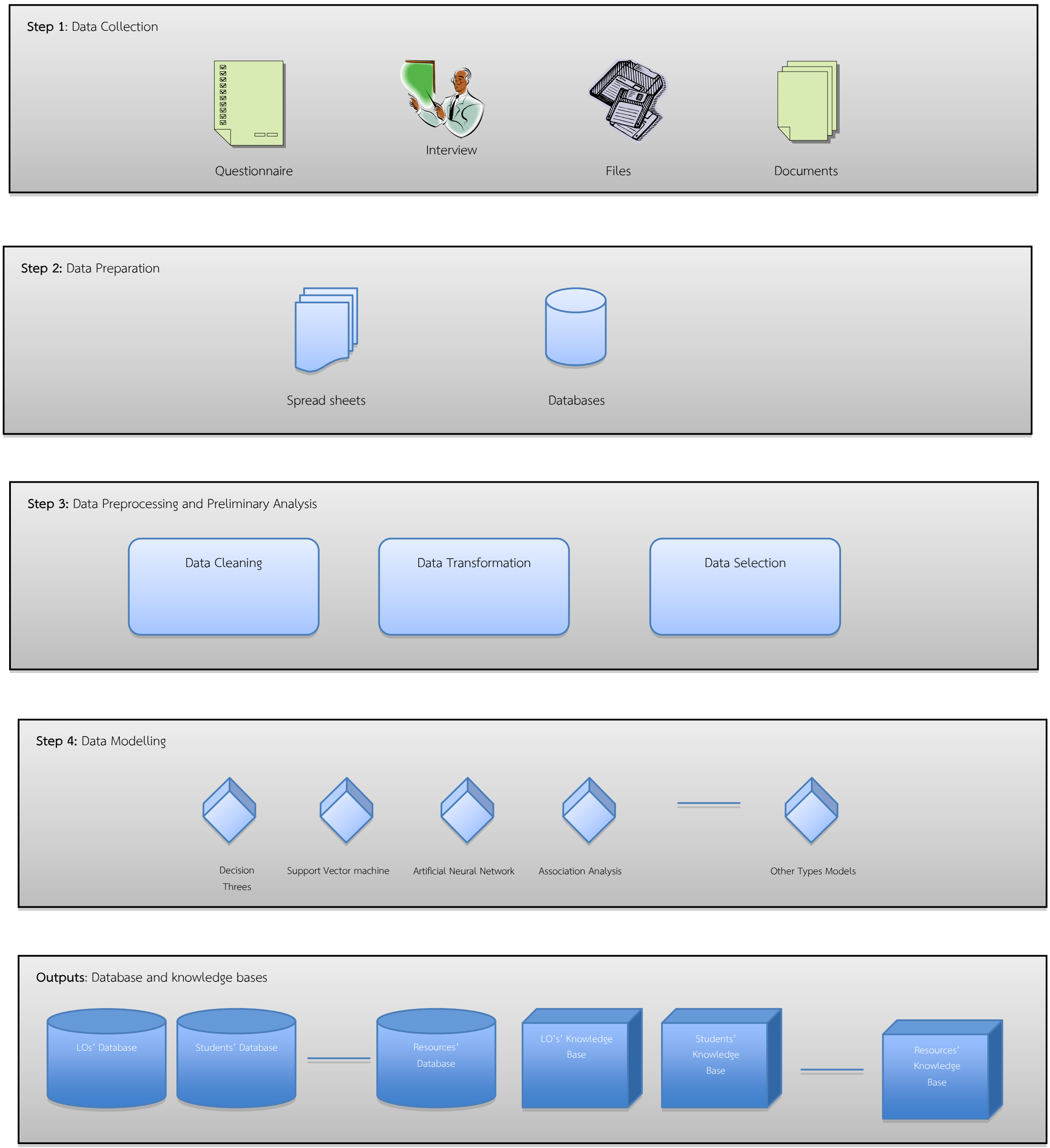

Figure 2 Data analysis process

Experiment Design

In figure 2, data mining process will be used. One of the modules from Kongsakun [19] can be employed in the classroom recommendation system is drop-out recommendation,

In this study, the proposed techniques have also been applied against the benchmark techniques; ANN, CHAID and SVM. In the proposed techniques, K-means clustering, an unsupervised learning is employed in order to classify student dropouts, the three benchmark techniques; SVM, CHAID and ANN are then applied in each cluster. All outputs of each cluster are compared and the two higher accuracy models are chosen, all clusters of the same models are combined before the steps of aggregation. In the combination process, two aggregation techniques; Ensemble method based on confidence-weight voting and Modular Artificial Neural Networks-Optimized Weight of Subspace Reconstruction Method (MANN-OWSR) are employed to aggregate two models that gives higher accuracy with the combination of all cluster of the same models. The outputs of these two aggregation models are compared, and the best accuracy model is then chosen. In the final step, the chosen model is again compared with the benchmark models. In the final step, the best model is chosen for the module of dropout identification in the intelligent recommendation system. In the proposed model, SVM, Neural Networks, CHAID and K-Means clustering algorithms are employed and the proposed model is shown in the diagram as shown in figure 3.

In the figure 3, it shows the processes of the student dropout identification model. The first process is the bench mark techniques which input data have trained, validated and tested by SVM, CHAID and ANN in three times of each techniques to ensure that the results are not obtained by chances. The results of these three techniques are compared with the final results of the proposed technique. The second process is applied by K-Means clustering. The survey results by interviewing the supervisors and counselors' show that the dropouts can classify for two main groups; low level of GPA and personal reasons; for example, family problem or financial problem. Therefore, the input data are clustered into two groups of related data. Each cluster is applied in the next process. Third process used the same techniques as benchmark, but these three techniques are employed to train, validate and test the data in each cluster and in three time training, validating and testing data. The results of each cluster in each model are compared in the 1st comparison stage. Then the models that give the highest accuracy in each cluster are combined by Ensemble in order to improve the accuracy. Similarly, the fifth process is the combinations of the models that give the second accuracy in each cluster using Ensemble in order to improve the accuracy. The results of these two processes are compared in the second comparison stage. Next process is applied by MANN-OWSR to aggregate the results of cluster 1 and cluster 2 of the highest accuracy model in the comparison of process 4 and 5. This process gives the final output of the proposed techniques. In the final comparison stage, the final results of the proposed techniques are compared with the results of the benchmark techniques. Then the model which gives the best results is chosen for the intelligent recommendation system. However, the results of these models are in binary type which is told that if students are high risk of dropouts, the results are "1" and if students can complete the degree, the results are "0". In regarding to the results shown in binary format, the metrics used to validate the results in this experiment is the accuracy percentage which is presented in the possibility table and graph format.

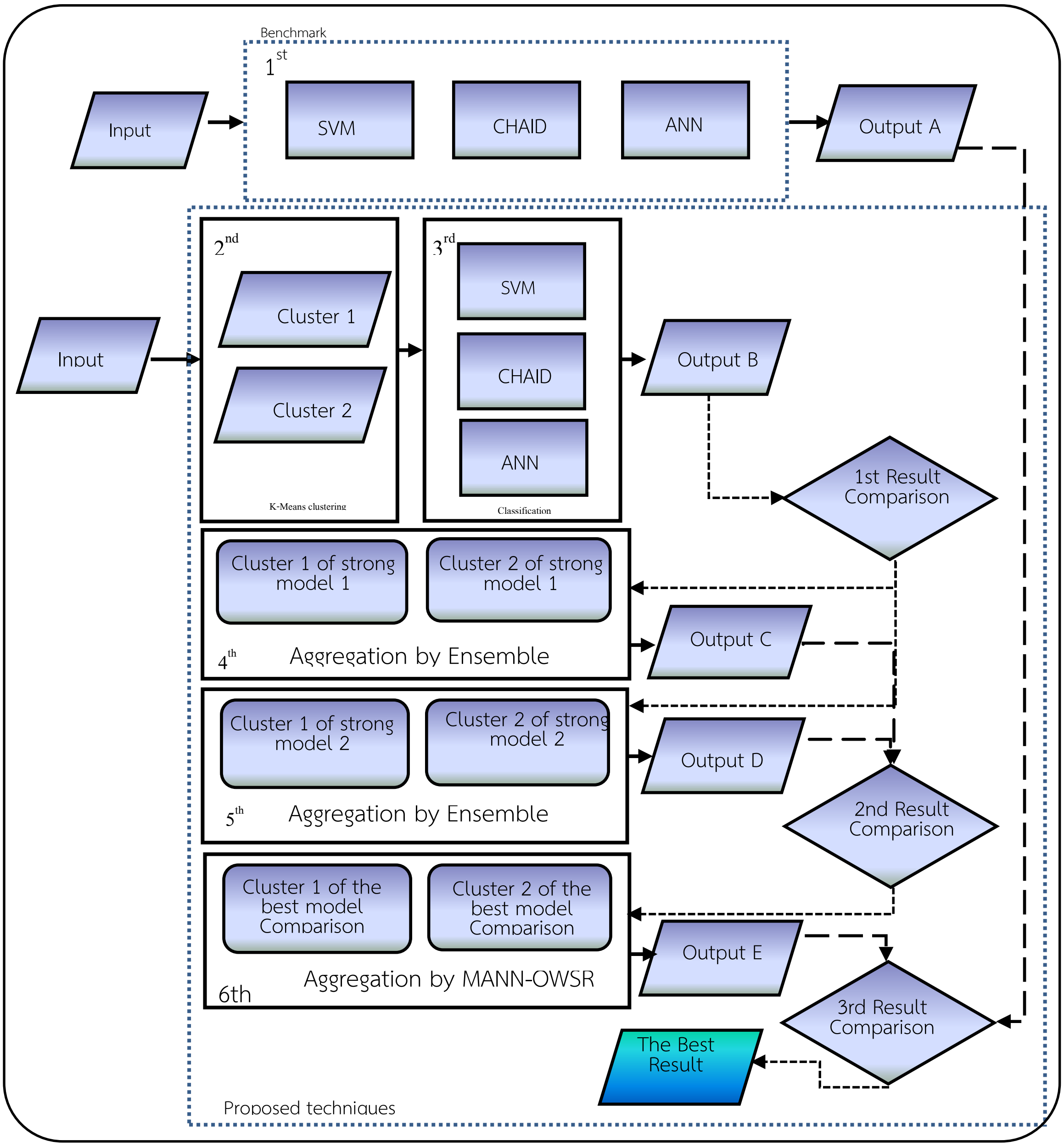

Figure 3 The student dropout identification model

Experimental Results of Drop-out recommendation

The research based on the experimental results, concentrates on the improvement of the prediction results in order to choose the best results for the intelligent recommendation system. The comparison of the accuracy rate of the recommendation system are shown in figure 4-9.

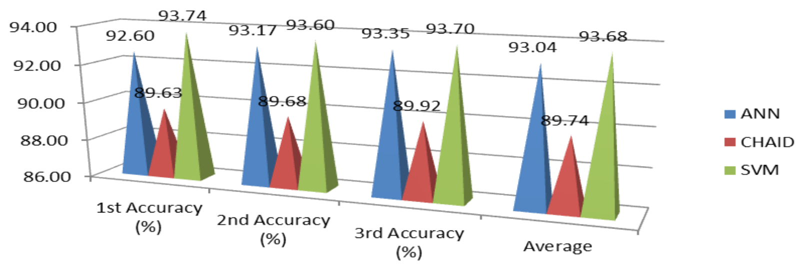

Figure 4 Accuracy of benchmark classification model; ANN, CHAID and SVM models in three time testing data

The comparison results of the benchmark techniques show that SVM model returned the highest accuracy of 93.68; therefore, SVM based model can be utilised to predict the student dropout identification with the best degree of accuracy, and this model is chosen to compare with the final results of proposed techniques in the last process.

In the second process, K-Means clustering is employed to separate data. This experiment used two clusters due to the survey results shown that the dropout can classify for two main groups; low level of GPA and personal reasons; for example, family problem or financial problem. Then, the results obtained that the number of cluster 1 and 2 are 4608 and 6792 records respectively. In the next section, the two clusters are applied by three techniques as shown below.

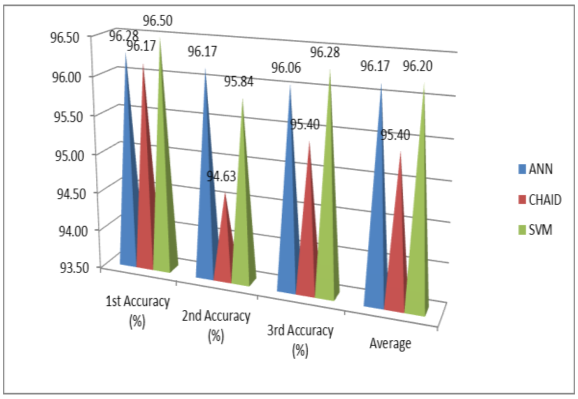

Figure 5 accuracy cluster 1 of three classification models in three time testing data

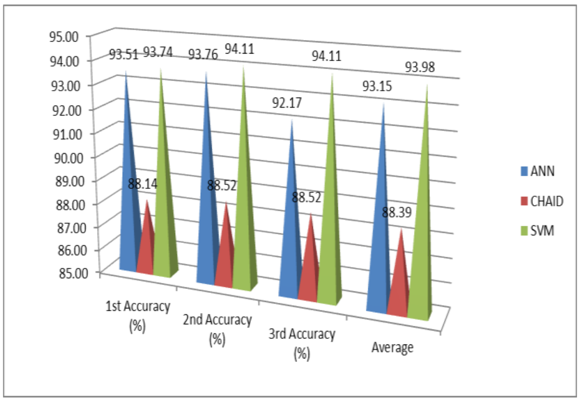

Figure 6 Accuracy of cluster 2 of three classification models in three time testing data

The comparison results demonstrate that SVM and ANN model return the higher accuracy than CHAID in both cluster 1 and 2; therefore, these 2 models are chosen to aggregate in the next process.

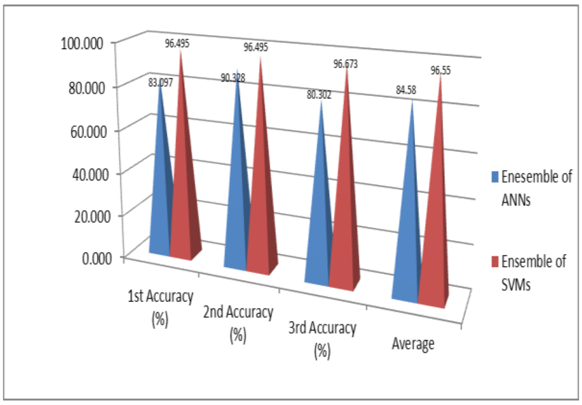

Figure 7 Accuracy of Ensemble of cluster 1 and 2 of SVM models in comparison with the combination of cluster 1 and 2 of ANN models method in three time testing data

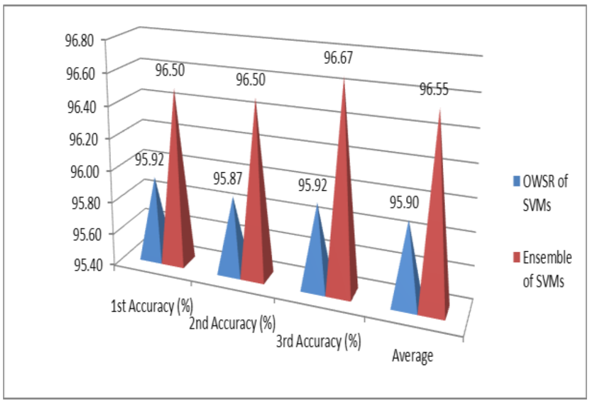

Figure 8 the accuracy of the Ensemble of cluster 1 and 2 of SVM models in comparison with the results of MANN-OWSR of cluster 1 and 2 of SVM models.

In figure 7, the comparison results of aggregation show that Ensemble of SVM combined cluster returned the highest accuracy of 96.55 whereas Ensemble of ANN return the accuracy of 84.58 percentage. Therefore, SVM clusters are chosen to aggregate using MANN-OWSR and Ensemble in the next process. In figure 8, the comparison results of Ensemble of SVM clusters return the highest accuracy of 96.55 percentages whereas the results of MANN-OWSR return 95.90 percentages.

It can be noted that Ensemble of SVM clusters is chosen to compare with the benchmark model in the last process as shown in figure 9.

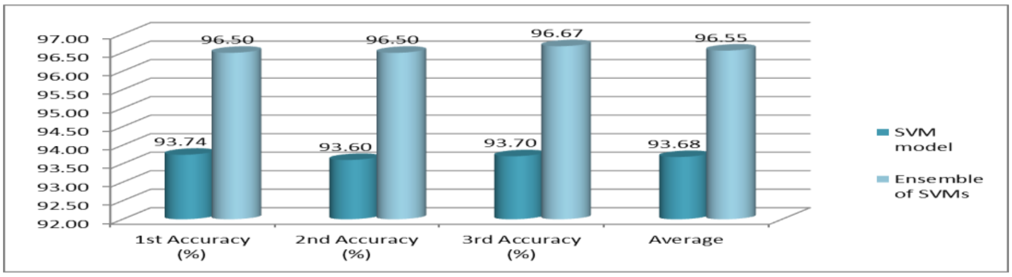

Figure 9 The accuracy of the Ensemble of cluster 1 and 2 of SVM models in comparison with the results of SVM model from benchmark techniques.

In figure 9, the accuracy of the Ensemble model of SVM cluster models in comparison with the results of SVM model from benchmark techniques. The results indicates that Ensemble of SVM models returned the highest classification accuracy of 96.55 percentages, while SVM models from the benchmark techniques returned the average accuracy 93.68 percentage. Therefore, Ensemble of SVM models outperforms the SVM models from the benchmark techniques.

Conclusions

In this experiment, drop-out module is employed. It is found that Ensemble of SVM models outperforms the SVM models from the benchmark techniques and every model which has tested with this dropout identification model. In this study, the experiment results show that Ensemble of SVM models can be utilised to predict the student dropout to identify students who are risk of leaving the study before graduation with the greatest degree of accuracy. Therefore, the best model with highest accuracy is chosen to use in the Intelligent Recommendation System for classroom. Next modules for recommendation will be experimented.

References

O. o. H. E. C. (2008). (2014, 30 January 2015). *Higher Education Institutions recognized by Office of the Higher Education Commission* Available: http://inter.mua.go.th/main2/index.php

J. J. McArdle, F. Hamagami, J. Y. Chang, and E. S. Hishinuma, "Longitudinal Dynamic Analyses of Depression and Acadmic Achievement in the Hawaiian High Schools Health Survey Using Comtemporary Latent Variable Change Models," *Structural Equation Modeling: A Multidisciplinary Journal,* vol. 21, pp. 608-629, 2014.

J. Kushner and G. M. Szirony, "School Counselor Involvement and College Degree Attainment: A Quantitative Conundrum," *Online Journal of New Horizons in Education,* vol. 4, pp. 75-79, 2014.

C. Vialardi, J. Bravo, L. Shafti, and A. Ortigosa, "Recommendation in higher education using data mining techniques," presented at the 2nd International Conference on Educaitional Data Mining, 2009.

K. Kongsakun, "An Intelligent Recommendation Framework for Student Counselling Management in Thai universities," Doctor of Information Technology, School of Engineering and Information Technology, Murdoch University, Research Repository, 2013.

B. Wu, L. Qi, and X. Feng, "Personalized Recommendation Algorithm based on SVM," presented at the International Conference on Communications, Circuits and Systems(ICCCAS2007), 2007.

S. Wanapu, C. C. Fung, N. Kerdprasop, N. Chamnongsri, and S. Niwattanakul, "An Investigation on the correlation of learner styles and learning objects characteristics in a proposed learning Objects Management Model (LOMM)," *Education and Information Technologies,* 2014.

M. Bienkowski, M. Feng, and B. Means, "Enhancing Teaching and Learning Through Educational Data Mining and Learning Analytics: An Issue Brief," in *office of Educational Technology*, U. S. D. o. Education, Ed., ed. Washington, D.C., U.S.A.: U.S. Department of Education, 2012.

E. Ozpolat and G. B. Akar, "Automatic Detection of Learning Styles for an eLearning System," *Computer and Education,* vol. 53, pp. 355-367, 2009.

F.-M. Shyu and H.-Y. Liao, "An application of SVM: Blog templates recommendation system in 2010 " in *Sixth International Conference on Natural Computation (ICNC 2010)*, 2010.

J. Gao and Z. Lishou, "Research on Intelligent Recommendation System of Teaching Based on web," in *2012 International Conference on Cybermatics and Informatics Lecture Notes in Electrical Engineering* 2012, pp. 2399-2404.

C.-M. Chen, "Recommending collaboratively generated knowledge," *Computer Science and Information Systems,* vol. 9, pp. 871-892, 2008.

C. F. Lin, Y.-C. Yeh, Y. H. Hung, and R. I. Chang, "Data Mining for providing a personalized learning path in creativity: An Application of Decision Trees," *Computer and Education,* vol. 68, pp. 199-210, 2013.

R. Kala, A. Shulkla, and R. Tiwari, "Fuzzy neuro systems for machine learning for large datasets," in *IEEE International Advance Computing Conference(IACC2009)*, Patiala, India, 2009.

S. Chabaa, A. Zerocal, and J. Antar, "Identification and prediction of Internet Traffic using artificial neural networks," *Journal of Intelligent Learning Systems and Applications,* vol. 2010, pp. 147-155, 2010.

D. Barbella, S. Benzaid, J. Christensen, B. Jackson, X. V. Qin, and D. Musicant. Understanding Support Vector Machine Classifications via a recommender System-Like Approach [Online]. Available: file:///Users/kanokwankongsakun/Google%20Drive/1.Applying_Scholarship-JOB_Australia/1.ResearchFellow_Dr.Kanokwan/svmzen.pdf

D. Martens, B. Baesens, and T. V. Gestel, "Decompositional Rule Extraction from Support Vector Machines by Active Learning," *IEEE Transaction on Knowledge and Data Engineering,* vol. 21, pp. 178-191, 2009.

M. M. Pozo, J. A. Iglesias, and A. I. Ledezma, "Intelligent Data Engineering and Automated Learning- IDEAL 2014," in *Lecture Notes in Computer Science* vol. 8669, ed, 2014, pp. 231-238.

K. Kongsakun, "Drop-Out Identification model using Data Mining for an Intelligent Recommendation System for universities in Thailand.," *Proceeding of the Hatyai Symposium 2013,* 2013.