A proposed framework in an intelligent recommender system for the college student

D Kurniadi1,2*, E Abdurachman1, H L H S Warnars1 and W Suparta1

1 Department of Computer Science, BINUS Graduate Program - Doctor of Computer Science, Bina Nusantara University, Jakarta, Indonesia
2 Department of Informatics, Sekolah Tinggi Teknologi Garut, Garut, Indonesia

*dede.kurniadi@binus.ac.id

Abstract. This article aims to proposed framework an Intelligent Recommender System (IRS) for students in higher education institutions. This conceptual framework includes problems in predicting student performance, the possibility of graduating on time, and recommends choosing subjects according to performance, and career interests, which are useful for assisting pedagogical interventions in future student development. The success in the development and implementation of the proposed IRS framework is inseparable from using data mining and machine learning techniques in predicting and providing recommendations. Data analysis consisted of clustering techniques, association rules, and classification using Support Vector Machine (SVM), Naïve Bayes, and k-Nearest Neighbour (k-NN). These techniques are used to solve problems related to students and to provide appropriate recommendations. The result is an IRS conceptual framework for the college student that can be used as smart agents to provide student guidance and suggestions to support the process of education in higher education.

1. Introduction

Academic information systems are a place of storage for data collection, processing, analysis and reporting of educational information. Besides aiming to collect, store data and process information, it also helps in monitoring, policy formulation, decision making, management, and evaluation. Large amounts of data stored in academic databases. This database contains information that is useful if explored more deeply to foster future students so that the risk of failure of studies can be avoided [1]. On the other hand, improving the quality of education in higher education can be seen from one of the high rates of student success and low failure rates of students. In Indonesia based on 2017 statistics, the number of students who experienced study failure was 195,176 people [2]. That has become a challenge for higher education institutions to maintain their reputation and business continuity.

John MC Carthy first introduced the term Artificial Intelligence (AI) in the field of computer science in 1956 and developed after Alan Turing proposed ideas about machines that can simulate humans and the ability to do smart things [3]. AI technology in the form of intelligent recommender systems (IRS) has also developed along with massive data growth in electronic form and is not only used in the domains of business, health, entertainment, government, but also the field of Education [4,5]. In the area of education, the IRS has an essential role in guiding students in the future. The IRS is an intelligent software application process analytical data using machine learning technology with the use of historical data and real-time in making predictions and recommendations automatically [6]. Analytic of data
education aims to study the data that comes from the area of education so that they can better understand the student's character and bring up the hidden knowledge [7]. The analytic process is the process of discovering, analysing, and interpreting meaningful patterns from large amounts of data [8]. Analysis of data mining technique relying on data such as classification, Association, correlation, estimation, prediction, categorization, grouping, analysing trends, and visualization [9]. So that the projections and recommendations given are not only based on expectations but come from existing empirical facts. Besides, the advantages of using data mining are rules or knowledge recommendations that can be updated based on new circumstances that may arise and influence the results.

In Indonesia, the education standard for Higher Education levels widely uses semester credit units (SCU). SCU is a system for implementing educational programs that frees students to determine for themselves how much learning burden will be followed in each semester [10,11]. With a system of SCU, students can plan courses of the higher semester based on their performance last semester. But time limitations and capabilities of the academic supervisor are traditionally determining barriers and fostering students following their interest. Students are not aware of the subjects chosen to set her career path so that it will have an impact on academic performance the next year. The selection of the appropriate subjects in the study period played a critical factor in shaping the future of the students, improving product performance timely graduation and academic.

Academic performance can be illustrated from close to learning outcomes at each course taken. To maximize the process of counselling for studying will help plan and understand the subject better way and can improve academic performance. It is possible to know the ability of students in learning courses last semester and also similar dug pattern learning from a previous academic database. So to produce the accumulation of excellent academic performance can be started by analysing and predicting performance based on each item subjects, then recommend the courses for next semester that match similar interest and there is nothing academic achievement of related classes [12–14]. This can be done by developing intelligent recommendation system as a smart agent specifically for college students, who can provide academic performance prediction, the likelihood of graduating on time, and give recommendations to choose appropriate subjects that match the achievement on the next semester without having to involve the academic supervisor directly.

Based on the background described, the purpose of this article is to propose an IRS framework that can be used as an intelligent agent of students to support the process of education in higher education. As for the focus of the discussion of the article is on student problems concerning the assessment of academic performance, the possibility of graduating on time and recommendations for choosing appropriate subjects that match the achievement of ongoing students or undergoing education at the university.

2. Methodology
The method used in writing this paper is based on the literature review of previous research. First introduced Educational Data Mining in the recommendation system, secondly identifying existing research problems and gaps related to student problems in the education domain, thirdly identifying machine learning and data mining methods used by previous researchers. The final stage is compiling the results of the study based on problem groups and method approaches so that there are opportunities to propose new frameworks with different methods.

3. Result and discussion

3.1. Educational data mining in recommendation system
Data Mining is the process of finding the model and relationships between data and is used to classify and analyse large databases [15]. Educational Data Mining (EDM) is a scientific discipline that is concerned with developing, researching, and applying computational methods to study the data that comes from the world of education so that they can better understand the characters of college students [16,17]. EDM is an exciting new field and rapidly developing that combines many disciplines include
computer science, science education, and statistics. This is also confirmed by Cristobal Romero and Sebastian Ventura in his research i.e. Data Mining Journal in Education that this research domain is an exciting research topic that continues to be developed one of the recommendations is developed Generic frameworks and methods to build the tools, frameworks, methods, algorithms, approaches, etc., specifically research-oriented Educational data mining [18]. While the recommendation system is the software tools and techniques that give suggestions for items that most likely to be of interest to a specific user [19]. Thus, the methods used in data mining Education became an essential component in building an intelligent recommendation system.

3.2. Problem studied and prediction method used
In the area of education, Intelligent Recommendation System has an essential role in guiding students in the future, such as predicting student performance based on prior knowledge, a recommendation of courses, student career prediction, prediction of graduation on time, the chances of dropout, until the projections of scholarship recipients. Conclusions based on the study of article 2014 – 2018 year related to the problem of the student includes student performance [20–22], course selection [23], graduation [24], dropout [25,26], career [27], and scholarship [28,29], and see issues regarding work student performance performed continuously as in figure 1.

Other work on the system recommendations for improving the academic performance of the students recently based on background factors (family and education before University entrance) was done by Goga et al. [30]. One of the limitations of the study in the collection of data used is limited to first-year students, which then the results of the work that had been done is difficult to differentiate between the accuracy of the framework developed and the accuracy of other frameworks. So further research is needed with a more effective method of approach.

Figure 1. Related work in student problems.

Figure 2 illustrates the algorithm is the most widely used in solving academic performance based on the highest accuracy with cumulative attribute Grade Point Average (GPA) is Decision tree 91%, K-Nearest Neighbour 83%, Support Vector Machine 80% Neural Networks 75%, and Naïve Bayes 75% [31–33].

3.3. Proposed framework intelligent recommender system for college student
Intelligent Recommender System (IRS) the proposed applying statistical methods, artificial intelligence, and data mining techniques that are different from previous research. The IRS plan can be implemented and used as an intelligent agent to give students’ academic performance and provides recommendations for how to choose appropriate subjects that match the achievement of students who are currently in progress or educated at University, so minimize the failure study in the middle of the street that had a significant impact on students' graduation. Jose Aguilar et al., has proposed a general framework for an intelligent recommendation system that expands the concept of knowledge-based recommendation system [6], as shown in figure 3.
There are four main elements as the difference of the recommendation of the knowledge-based approach in the IRS namely: Knowledge Acquisition, Knowledge Modeling, Reasoning, and Recommendation. Knowledge acquisition is a mechanism based on machine learning algorithms; knowledge modeling is explicit knowledge, which represents all the knowledge that is needed for recommended. Meanwhile, Reasoning is verification of the mechanism of thinking to retrieve information from the knowledge stored, and Recommendation whereas a system based on the inquiries automatic inference capabilities the IRS.

Figure 3. General architecture for intelligent recommendation systems.

Figure 4 illustrates a proposed framework of intelligent recommender system (IRS) for the college student. This proposal aims to analyse student based on academic record, the results close to the item the previous subjects, and the score of each subject every semester and an overall GPA of academic databases. The results will then be used to predict academic performance and graduation. Recommendation system will then provide recommendations for the most appropriate subjects for students following academic achievement.

Figure 4. The proposed framework IRS for the college student.

The proposed framework uses data mining and machine learning techniques consisting of Clustering, Association Rule, and Classification techniques using support vector machine, Naïve Bayes, and k-Nearest Neighbour, the selection of the combination of methods was chosen based on the results of the literature review from previous studies. The stages include of Storing Data, Data Analysis, and Visualization with Predictive Modelling, with the following steps.

3.4. Data storing
This step is the preparation of the data collected from an academic database. The data is then reformatted at the stage of data transformation to prepare the processing with the next algorithm. In the process of
cleaning the data, the parameters used in the analysis of data is identified, and the missing data are discarded, then separated into the training data.

3.5. Analysing data
In the stages of data analysis, Clustering techniques with K-Means algorithm are used to classify data of the same objects based on historical academic performance [34]. From the perspective of machine learning, the cluster corresponding to the hidden pattern and cluster search categorized as learning without supervision. Furthermore, the Association Rule used to examine the relationship between the subgroup by identifying the features most frequently appearing students based on the number of instances that are predicted correctly [35]. The most common outputs will be analysed and compared to clustering output to ensure more correct predictions. Next is the generation model, where we compare the performance of three classifiers namely, Support Vector Machine, Naïve Bayes, and k-Nearest Neighbour. The best performing classifiers will then be used to predict and classify academic performance data, the possibility of graduating on time, and recommendations for subjects.

3.6. Visualization
At this stage is to describe future prediction and recommendation results of data analysis and interpretation of prediction models with output in the form of predictions of possible GPA obtained for year 1 to year 4, forecasts of the possibility of graduating on time, and recommending subjects according to performance in the next semester.

4. Conclusion
The proposal of an intelligent recommender system (IRS) framework for college students in this article is to provide and build an IRS system in the form of a smart agent that can be used as a guide for students in the future. The focus of the next work is on testing, developing, and implementing the proposed framework, including the process of identifying parameters to be used, testing models, and recommendations for solving problems faced by students who are undergoing education in college.

References
[1] Alharbi Z, Cornford J, Dolder L and De La Iglesia B 2016 Using data mining techniques to predict students at risk of poor performance 2016 SAI Computing Conference (SAI) (IEEE) pp 523–531
[2] Ministry of Research Technology and Higher Education of Republic of Indonesia 2017 Higher Education Statistical 2017 Higher Education Statistical Year Book 1–194
[3] Flasiński M 2016 History of Artificial Intelligence Introduction to Artificial Intelligence (Cham: Springer International Publishing) pp 3–13
[4] Lu J, Wu D, Mao M, Wang W and Zhang G 2015 Recommender system application developments: A survey Decis. Support Syst. 74 12–32
[5] Kurniadi D, Mulyani A, Septiana Y and Aulawi H 2018 Estimated software measurement base on use case for online admission system IOP Conf. Ser. Mater. Sci. Eng. 434 012062
[6] Aguilar J, Valdiviezo-Díaz P and Riofrio G 2017 A general framework for intelligent recommender systems Appl. Comput. Informatics 13 147–160
[7] Bogarín A, Cerezo R and Romero C 2018 A survey on educational process mining Wiley Interdiscip. Rev. Data Min. Knowl. Discov. 8 e1230
[8] Berland M, Baker R S and Blikstein P 2014 Educational Data Mining and Learning Analytics: Applications to Constructionist Research Technol. Knowl. Learn. 19 205–220
[9] Kharade B and Wagh K 2016 Data Analytics in Educational Management System Int. J. Comput. Appl. 22–25
[10] Ministry of Research Technology and Higher Education of Republic of Indonesia 2015 National Standards for Higher Education
[11] Kurniadi D, Sasmoko, Warnars H L H S and Gaol F L 2017 Software size measurement of student
information terminal with use case point 2017 IEEE International Conference on Cybernetics and Computational Intelligence (CyberneticsCom) (IEEE) 164–169

[12] Kardan A A, Sadeghi H, Ghidary S S and Sani M R F 2013 Prediction of student course selection in online higher education institutes using neural network Comput. Educ. 65 1–11

[13] Ognjanovic I, Gasevic D and Dawson S 2016 Using institutional data to predict student course selections in higher education Internet High. Educ. 29 49–62

[14] Sobecki J 2014 Comparison of Selected Swarm Intelligence Algorithms in Student Courses Recommendation Application Int. J. Softw. Eng. Knowl. Eng. 24 91–109

[15] Xindong Wu, Xingquan Zhu, Gong-Qing Wu and Wei Ding 2014 Data mining with big data IEEE Trans. Knowl. Data Eng. 26 97–107

[16] Koedinger K R, D’Mello S, Melaughlin E A, Pardos Z A and Rosé C P 2015 Data mining and education Wiley Interdiscip. Rev. Cogn. Sci.

[17] Mishra A, Bansal R and Singh S N 2017 Educational data mining and learning analysis 2017 7th International Conference on Cloud Computing, Data Science & Engineering - Confluence (IEEE) pp 491–504

[18] Romero C and Ventura S 2013 Data mining in education Wiley Interdiscip. Rev. Data Min. Knowl. Discov. 3 12–27

[19] Ricci F, Rokach L and Shapiro B 2015 Recommender Systems: Introduction and Challenges Recommender Systems Handbook (Boston, MA: Springer US) pp 1–34

[20] Shahiri A M, Husain W and Rashid N A 2015 A Review on Predicting Student’s Performance Using Data Mining Techniques Procedia Computer Science 414–422

[21] Abu A 2016 Educational Data Mining & Students Performance Prediction Int. J. Adv. Comput. Sci. Appl. 7 212–220

[22] Al-Shehri H, Al-Qarni A, Al-Saati L, Batoaq A, Badukhen H, Alrashed S, Alhiyaffi J and Olatunji S O 2017 Student performance prediction using Support Vector Machine and K-Nearest Neighbor Electrical and Computer Engineering (CCECE), 2017 IEEE 30th Canadian Conference on 1–4

[23] Lin J, Pu H, Li Y and Lian J 2018 Intelligent Recommendation System for Course Selection in Smart Education Computer Science 129 449–453

[24] Ojha T, Heileman G L, Martinez-Ramon M and Slim A 2017 Prediction of graduation delay based on student performance 2017 International Joint Conference on Neural Networks (IJCNN) (IEEE) 3454–3460

[25] Amaya Y, Barrientos E and Heredia D 2015 Student Dropout Predictive Model Using Data Mining Techniques IEEE Lat. Am. Trans. 13 3127–3134

[26] Sivakumar S, Venkataraman S and Selvaraj R 2016 Predictive modeling of student dropout indicators in educational data mining using improved decision tree Indian J. Sci. Technol. 9 1–5

[27] Campagni R, Merlini D, Sprugnoli R and Verri M C 2015 Data mining models for student careers Expert Syst. Appl. 42 5508–5521

[28] Alhassan J K and Lawal S A 2015 Using data mining technique for scholarship disbursement Int. J. Inf. Commun. Eng. 9 1511–1514

[29] Kurniadi D, Abdurachman E, Warnars H L H S and Suparta W 2018 The prediction of scholarship recipients in higher education using k-Nearest neighbor algorithm IOP Conf. Ser. Mater. Sci. Eng. 434 012039

[30] Goga M, Kuyoro S and Goga N 2015 A Recommender for Improving the Student Academic Performance Procedia - Social and Behavioral Sciences vol 180 (Elsevier B.V.) 1481–1488

[31] Jishan S T, Rashu R I, Haque N and Rahman R M 2015 Improving accuracy of students’ final grade prediction model using optimal equal width binning and synthetic minority oversampling technique Decis. Anal. 2 1–25

[32] Natek S and Zwilling M 2014 Student data mining solution–knowledge management system related to higher education institutions Expert Syst. Appl. 41 6400–6407
[33] Mayilvaganan M and Kalpanadevi D 2015 Comparison of classification techniques for predicting the cognitive skill of students in education environment 2014 IEEE Int. Conf. Comput. Intell. Comput. Res. IEEE ICCIC 2014 113–118

[34] Durairaj M and Vijitha C 2014 Educational Data mining for Prediction of Student Performance Using Clustering Algorithms Int. J. Comput. Sci. Inf. Technol. 5 5987–5991

[35] Witten I H, Frank E, Hall M A and Pal C J 2016 Data Mining: Practical machine learning tools and techniques (Morgan Kaufmann)