Recommender System For High School Selection Based On Apriori Method

Herman Yuliansyah, Desita Putri Niranda, Ika Arfiani

Abstract: High school selection is a process to choose the best schools that appropriate for prospective students based on their competence and skills. Most prospective students cannot determine the appropriate school selection because students need guidance and consult with their Guidance and Counseling teacher to get advice. Later, most researchers focus on the implementation of data mining to analyze the student data of high education, namely university or polytechnic, so this research conduct to find for previous data patterns. The main objective of this research is to propose a recommender system to select the appropriate schools based on the apriori method in data mining. The recommender system considers previous knowledge from the dataset and extracts the dataset to be training data, then using the apriori method to determine the schools' selection. The recommender system is implemented in a web-based system. The web-based system evaluates using the system usability scale to get the user response based on user experiences and user-friendliness.

Index Terms: Recommender system, educational data mining, high school selection, association rules mining, apriori method.

————————————————   ◆   ————————————————

1. INTRODUCTION

SCHOOLS selection is a critical phase that affects the future of the prospective student. The education purpose is to develop the potential of students through the learning process under teacher supervision. Furthermore, students must register in the formal education system, which is compulsory. At present, the known formal education system includes primary school education, junior high school education, senior high school and the higher education system. Then, senior high schools consist of senior high schools and vocational high schools. The selection of the right school is an important issue to ensure a prospective student can register in the right school following the character of the prospective student's competence. Based on the results of a survey conducted for junior high school students, 60% of students are still unable to determine in the selection of high school. This phenomenon is because students have not been able to decide which school is following their potential whereas the selection of the right school following the potential of students can support the ability of students' development in the future. So students need to discuss asking for guidance from the Guidance and Counseling teacher. Guidance and Counseling teachers guide students by looking at the academic grades of grades 7,8,9, non-academic values include the achievements of these students. However, the recommendations given are not made to all students but only to students who ask BK teachers to get recommendations or directions for the selection of high school level. This phenomenon has the potential to develop a recommendation system because it is a tool for selecting schools. In previous research, we have conducted a study on data mining implementation to analyse alumni data[1] and student data[2][3] in high education, in addition to analysing health domain data[4][5][6]. Data mining in education has an essential impact to extract meaningful information[7][8] and improve the quality of learning process[9]. Later, association rules mining as part of data mining techniques also have wide

————————————————

- *Herman Yuliansyah, Informatics Departments, Universitas Ahmad Dahlan, Indonesia. E-mail: herman.yuliansyah@tif.uad.ac.id (Corresponding Author)*
- *Desita Putri Niranda, Alumni Informatics Departments, Universitas Ahmad Dahlan, Indonesia. Email: desita1500018190@webmail.uad.ac.id*
- *Ika Arfiani, Informatics Departments, Universitas Ahmad Dahlan, Indonesia. Email: ika.arfiani@tif.uad.ac.id*

implemented to analyse high schools students. Reff [10] propose course recommender system to predict the best courses combination that selected by students. Furthermore, student academic success can evaluate by predicting the final placement of student based on mark score of the previous semester. Moreover, data mining also used to analyse the relationships among student's behavioural and student success to develop the model of student performance predictors[11]. Another student performance is used to analyse the performance of undergraduate students study[9]. The analysed results are to prevent the risk of the student to get drop-out[12] then reduce the drop out ratio to better levels[13]. Association rules mining, clustering and classification, are the popular technique in data mining[14]. Most of the researcher focus on data mining for high education by analyzing inside the academic databases nor outside the academic[15]. So this research was conducted to search for patterns of recommendation for the selection of senior high schools using the association method and then implemented into a web-based system. The web-based system developed using python programming and MySQL database. Later, the junior high school and Guidance and Counseling teacher are the users of the web-based system. The recommendation consider several aspects, namely academic achievements and non-academic, which includes achievements in the fields of religion, sports, and science. This article organised in four sections: Section 2 is the proposed recommender system, section 3 is methodology research, section 3 is result and discussion based on experiment results, and section 4 is conclusions.

2 PROPOSED RECOMMENDER SYSTEM

The proposed recommender system consist of several part,
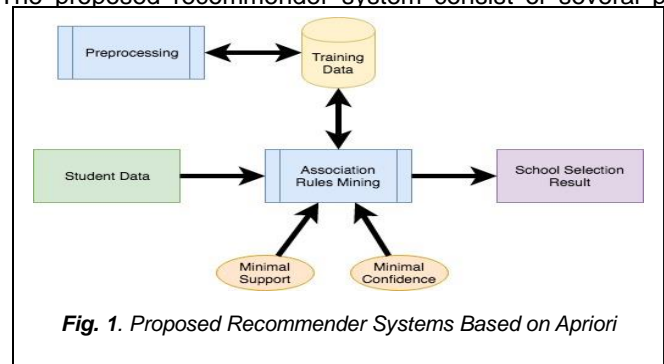
Fig. 1. *Proposed Recommender Systems Based on Apriori*

2360

namely probe data or input data, data training, recommender and recommender result as shown in Fig 1. The training data is student data history that collected from previous student and extract the information to get the pattern. The extraction of training data processed using preprocessing techniques to prevent the validity of data. The training data is very essential because would be as a guideline for recommender to determine the decision. The classifier use apriori method of association rules mining, which need two parameter value, namely minimal support and minimal confidence. The minimal support is a cut-off value to select the data will used or not. Later, the minimal confidence is a confidence level of apriori result. The user entry several value that correlate with to training data as a testing data. Furthermore, the recommender result is the school selection based on the apriori method result. There are two type result of this recommender, namely the option for senior high schools or vocational high schools.

3   METHODOLOGY RESEARCH

This research conducted with few step, as shown in Fig 2:

Fig. 2. Methodology Research
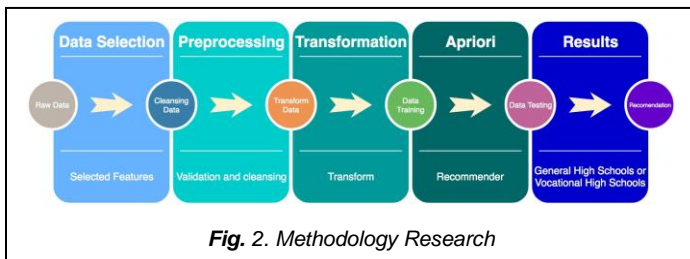
3.1 Data Selection
The data selection stage is carried out before the stage of extracting information. The selected data will be saved in a separate file for further use.

3.2 Preprocessing
The preprocessing stage is carried out to remove or clean invalid data, lost data or misprinted data. data selection can affect the performance of a data mining system because the data used is decreasing or less because only relevant data is left.

3.3 Data Transformation
The transformation phase is carried out to transform forms that are not yet suitable into valid data forms for further data mining processes.

3.4 Apriori Method
This stage applies an algorithm or some desired method to search for patterns or information in selected data. The choice of the right method is very dependent on the goals and process of data mining.

3.5 Evaluation
The last step taken is to display the results of the process of forming data mining output in a form that is more easily understood by other users. This evaluation phase is also useful for checking the information obtained, whether or not it matches the previous facts.

4   RESULTS AND DISCUSSIONS
The recommender system process divided into five stages: 1)

Load Data, 2) Data Selection, 3) Data Transformation, 4) Frequent Itemset and 5) Apriori Results.

4.1 Dataset
The dataset from a private junior high schools that consist of 45 student data with variables, i.e., academic achievements and non-academic which includes achievements in the fields of religion, sports, and science, as shown in Tabel 1.

No	Student ID	Sub0	Sub1	Sub2	Sub3	Sub4	Sub5	Sub6	SubX	Schools
TABLE 1 *Dataset*										
1.	05-173-001-8	…	82	86	84	86	83	81	…	SMK
2.	05-173-002-7	…	82	86	82	83	82	81	…	SMK
3.	05-173-003-6	…	88	88	84	89	90	85	…	SMA
…	…	…	…	…	…	…	…	…	…	…
44.	05-173-044-5	…	93	94	98	99	95	85	…	SMA
45.	05-173-045-4	…	85	89	87	89	85	91	…	SMA

Note: Sub0 until SubX means academic achievements and non-academic value, SMK means vocational high schools and SMA means senior high schools or non-vocational high schools.

4.2 Load Data
This research not only analyses association rules of the student data based on apriori method but also develop web-

Fig. 3. Load spreadsheet
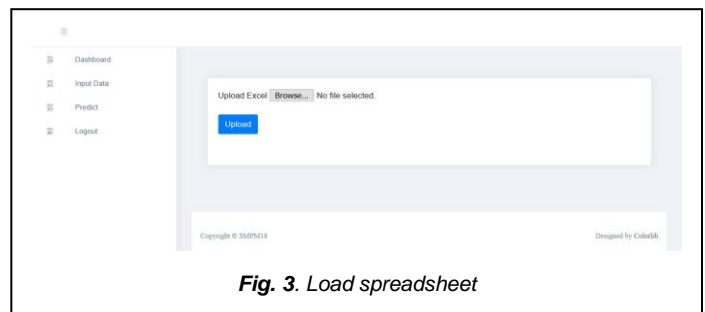
based system. After the data from Table-I store in spreadsheet file, then the user must upload the data into web-based system, as shown in Figure 3. The spreadsheet file load and read all of the data to preprocessing and ready to analyse into web-based system, as shown in Fig. 4. There are two processes in load data, i.e., load the spreadsheet, as shown in Fig. 3 and read and store the spreadsheet, as shown in Fig. 4. The load spreadsheet do upload file in csv extension into server, then the web-based system read column by column using semicolon delimiter and read row by row using new line format. After upload process finish, the file uploaded read from temporary storage. If the user feels the data value is valid. Then store the data into databases. The web-based system is designed using mix-language because to increase usability of web-based system, as shown in Fig. 4. AGAMA7 is religion subject for class 7th, B.IND7 is Indonesian language subject for class 7th, MTK7 is mathematic for subject for class 7th is Indonesian language subject for class 7th, IPA is natural science language subject for class 7th, IPS is social science subject for class 7th, B.ING is English subject for class 7th, and etc.

4.3 Data Selection

Fig. 4. Read the data and store in database

The data selection step (or Data Seleksi in web-based system) is a process to select or determine the used variable of stored previous data. This research only select variable such as value for several subjects of class 7th, 8th, and 9th. The subjects are religion, Indonesian language, mathematic, natural science, sosial science, English and non-academic value such as achievements in the religious, sports, scientific, national exam, and school final exam, as shown in Fig. 5.

Fig. 5. Data selection

Fig. 6. Data Transformation

4.4 Data Transformation
The data transformation step (or Data Transformasi in web-based system) is a process to change the format of data variable to formatted data. In the apriori method, the data will

Fig. 7. Frequent Itemset

be arranged in series and if there is the same data in the data arrangement, even though the variables are different, then the apriori method will read into the same data. This situation will affect the results rather than apriori methods and to overcome this it is necessary to transform the data by combining the variable name and value prefixes into a new value. This method is a solution so that the a priori method can read all the values of different variables, as shown in Fig. 6.

4.5 Frequent Itemset
Frequent itemset is the trial process to determine the minimum support value and based on our experiment, the minimum support value set to 5. The minimum support value mean the data of every variable that have support more than or equal than 5 will be used for the apriori method process, as shown in Fig. 7. The itemset that have equal or more that 5 counted the number of equal pattern itemset, as shown in Fig. 8. Frequent itemset process is an essential process because, the frequent itemset results produce number of itemset that use for next apriori process. Setting a minimum support value that is too low will result in many itemset that will appear and setting a minimum support value that is too high will reduce the existing itemset so that the minimum support value must be adjusted to the desired number of patterns to be achieved.

4.6 Apriori Results
Apriori method processed in background process of web-based system, and only the result shown in thw web-based system, as shown in Fig. 9. The users of web-based system are Indonesian people, so the results or the apriori results shown in Indonesian language. The results composed in simple sentence using IF THEN pattern. In the last apriori results shows "jika nilai B.IND kelas 8 adalah A, nilai IPA kelas 9 adalah A, nilai B.IND kelas 9 adalah A, nilai MTK kelas 9 adalah A, maka rekomendasinya adalah SMA dengan konfidensi 95%" means "If value subject of Indonesian language class 8th is A, value subject of natural science class 9th is A, value subject of Indonesian language class 9th is A, value subject of Mathematics class 9th is A, then the recommendation is Senior High Schools with confidence value is 96%".

4.7 Predicts
The student or guidance and counseling teacher is the user of these web-based system. The user can find the recommendation of a student achievement based on the pattern that has contruct from apriori methods. The user only fill several variable, such as student name (Nama Siswa), student number (NIS), value subject of class 7th, 8th, 9th (nilai kelas 7, nilai kelas 8, nilai kelas 9), religion achievement (prestasi keagamaan), sports achievement (prestasi keolahragaan), science achievement (prestasi keilmuan) and national exam (Nilai UN), as shown in Fig. 10.

2362

Fig. 8. *Minimal Support Frequent Itemset*
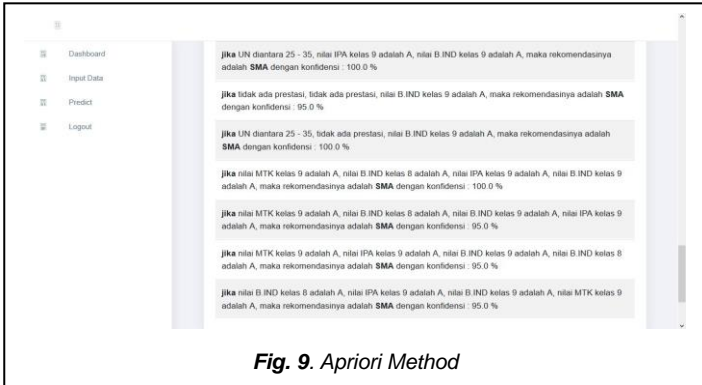

Fig. 9. *Apriori Method*

The users must fill all of the required fields to get the result of web-based system and push button "Prediksi". There is two-recommendation result of the web-based system, namely senior high schools (SMA) or vocational high schools (SMK), as shown in Fig. 11.

4.8 Software Usability Scale
Software usability scale (SUS)[16] questionnaire is used to measure the readiness of this web-based system. SUS is likert scale and contains ten questions. Each question consist of 5 option with 0-4 values and for question number 1,3,5,7, and 9
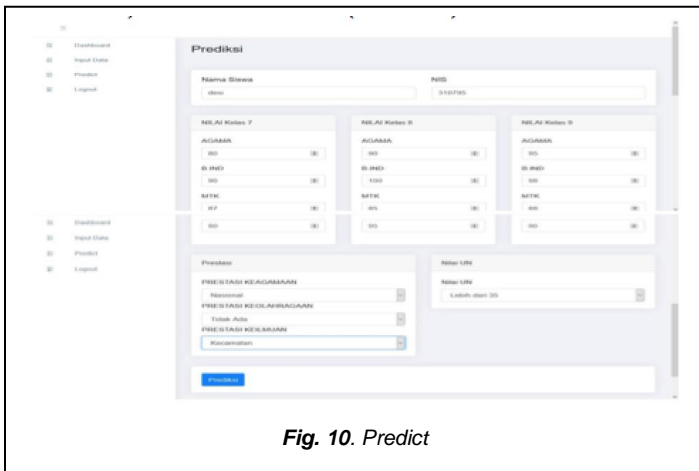

Fig. 10. *Predict*

are positif question with final score minus 1. Later, question number 2,4,6,8, and 10 are negative question with final score 5 minus the filled score. Furthermore, mutiply the sum of all score by 2.5. The final SUS score is average SUS score of every responden, as shown in Tabel 2The SUS result of this web-based system shows 74 score. This 74 score means "Good" for adjective ratings, "C" for grade scale and "Acceptable" for acceptability ranges.

5 CONCLUSIONS
Based on the results of the study can be concluded as follows:

Fig. 11. *Recommendation result (Hasil Prediksi)*

1. The development of a recommendation system using apriori methods developed with the results obtained in the form of rules, which are a collection of frequent itemsets with high confidence values. With these rules obtained, guidance and counselling teachers can use these rules in providing overall high school recommendations to students.
2. Implementation of apriori methods from the 45 data history of junior high school students the minimum support requirement is five and resulting in 33 rules that recommend senior high schools.
3. The web-based system is "Acceptable" based on Software usability scale that means the system ready to use.

No	R1	R2	R3	R4	R5
			TABLE 1 *Dataset*		
1	4	3	3	4	4
2	1	1	3	1	1
3	2	3	1	2	1
4	3	3	2	1	4
5	3	3	3	4	4
6	3	2	3	3	3
7	4	3	4	4	4
8	4	3	4	4	4
9	3	3	2	4	4
10	3	3	4	2	4
Jml	30	27	29	29	33
X 2,5	75	67,5	72,5	72,5	82,5
Rata – rata = 370/5 = 74					

REFERENCES
[1] H. Yuliansyah and L. Zahrotun, "Designing web-based data mining applications to analyze the association

2363

rules tracer study at university using a FOLD-growth method," Int. J. Adv. Comput. Res., vol. 6, no. 27, pp. 215–221, Oct. 2016.

[2]   K. Sya'iyah, H. Yuliansyah, and I. Arfiani, "Clustering Student Data Based On K-Means Algorithms," Int. J. Sci. Technol. Res., vol. 8, no. 8, pp. 1014–1018, 2019.

[3]   H. Yuliansyah, Hafsah, I. Arfiani, and R. Umar, "Discovering Meaningful Pattern of Undergraduate Students Data using Association Rules Mining," in 2019 Ahmad Dahlan International Conference Series on Engineering and Science (ADICS-ES 2019), 2019, pp. 13–17.

[4]   I. Riadi, S. Winiarti, and H. Yuliansyah, "Development and evaluation of android based notification system to determine patient's medicine for pharmaceutical clinic," in 2017 4th International Conference on Electrical Engineering, Computer Science and Informatics (EECSI), 2017, no. September, pp. 1–5.

[5]   S. Winiarti, H. Yuliansyah, and A. A. Purnama, "Identification of Toddlers' Nutritional Status using Data Mining Approach," Int. J. Adv. Comput. Sci. Appl., vol. 9, no. 1, pp. 164–169, 2018.

[6]   S. Winiarti, S. Kusumadewi, I. Muhimmah, and H. Yuliansyah, "Determining the nutrition of patient based on food packaging product using fuzzy C means algorithm," in 2017 4th International Conference on Electrical Engineering, Computer Science and Informatics (EECSI), 2017, vol. 2017-Decem, pp. 1–6.

[7]   B. R.B, "'Importance of Data Mining in Higher Education System,'" IOSR J. Humanit. Soc. Sci., 2013.

[8]   K. Kohli and S. Birla, "Data Mining on Student Database to Improve Future Performance," Int. J. Comput. Appl., 2016.

[9]   R. Asif, A. Merceron, S. A. Ali, and N. G. Haider, "Analyzing undergraduate students' performance using educational data mining," Comput. Educ., 2017.

[10]   S. B.Aher and L. L.M.R.J., "Prediction of Course Selection by Student using Combination of Data Mining Algorithms in E-Learning," Int. J. Comput. Appl., 2012.

[11]   S. Sembiring, M. Zarlis, D. Hartama, S. Ramliana, and E. Wani, "Prediction of Student Academic Performance By an Application of Data Mining Techniques.," 2011 Int. Conf. Manag. Artif. Intell., 2011.

[12]   P. Paruechanon and W. Sriurai, "Applying association rule to risk analysis for student-dropout in Information Technology Program," J. Sci. Sci. Educ., vol. 1, no. 2, pp. 123–133, 2018.

[13]   H. Islam and M. Haque, "An Approach of Improving Student's Academic Performance by using K-means clustering algorithm and Decision tree," Int. J. Adv. Comput. Sci. Appl., 2013.

[14]   J. Han, M. Kamber, and J. Pei, Data Mining: Concepts and Techniques, 3rd ed. USA: Morgan Kaufmann, 2012.

[15]   M. Wati, W. Indrawan, J. A. Widians, and N. Puspitasari, "Data mining for predicting students' learning result," in Proceedings of the 2017 4th International Conference on Computer Applications and Information Processing Technology, CAIPT 2017, 2018.

[16]   J. Brooke, "SUS-A quick and dirty usability scale," Usability Eval. Ind., vol. 189, no. 194, pp. 4–7, 1996.