Pattern Capacity Participants Exam Of Mobile Learning For Assessment

Heri Nurdiyanto, Herman Dwi Surjono, Priyanto

Abstract: The adaptive assessment system is expressed as an interactive approach to assessing the learner in the learning system. Stages undertaken in the development of this system include determination of bank questions, determination of the initial ability level of examinees, selection of items, assessments, termination of tests, and conclusions about the ability of examinees. Determining the initial ability level of examinees is very important because its accuracy dramatically affects the effectiveness of a selection of questions. Rule-based methods are used to extract information, rule-based methods combined with machine learning techniques are proposed to assess the level of ability of regular students and students with special needs. Machine learning techniques used are Naive Bayes, Multilayer Perceptron, SMO, Decision Tree, JRIP, and J48. The best accuracy results are achieved using the JRIP rule-based method of 64.12. The rules for the determination of the level of ability are formed based on expert opinion. The strength of examinees to vary and the amount of data evolving lies in need for dynamic formation of rules. The discovery of patterns in the test data of the participants can be used as the basis for the creation of states to replace the expert as well as improve the prediction accuracy. It is necessary to extract the pattern so that it can be used for the formation of the initial capability rules for examinees.

Keywords: K-means, a rules-based classification method, scoring system.

1. Introduction

Currently, information technology has been widely used to help the learning process. At the high school level, web-based e-learning is built to meet the needs of schools[1]. In fact, user-based e-learning has been developed to provide learning materials that match the student's learning styles as well as the user's conditions while learning[2]. In addition to helping the learning process, information technology is also applied to the assessment of learning. Evaluation of education is the process of evaluating the achievement of knowledge, understanding, and learning skills [3]. Assessment can be used for several needs, such as entrance exams, the basic design of learning materials, competency measurement, learning achievement measurement, and graduation determination[4]. The computer-based scoring system began to be widely used. Computer-Based National Exam is suitable for entrance exams, designing learning materials, and graduation exams[5]. However, to measure the competence or achievement of learning, it is necessary a system that can choose and present the problems by the ability of examinees. Adaptive scoring systems can be used to meet this need as these systems can select and give the questions according to the testers' knowledge[6]. The adaptive assessment system is expressed as an interactive approach to assessing learners in the learning system. The steps undertaken in the development of this system include determination of problem banks, determination of initial ability level of examinees, selection of questions, assessment, cessation of tests, and conclusions about the ability of examinees [7].

- *Departement Technology and Vovational Education, Yogyakarta State University, Indonesia*
- *2 & 3Faculty of Engineering, Yogyakarta State University, Indonesia*
- *Corresponding author's Email: heri.nurdiyanto2016@student.uny.ac.id*

Determining the initial ability level of examinees is very important because its accuracy dramatically affects the effectiveness of a selection of questions. Adaptive valuation system has been developed. However, there is little focus on determining the level of examinees' ability[8]. The ability level of examinees can be classified into several categories, among others, based on the concept being studied, the competencies held, the bloom rate, and the classes being pursued. In one study, the Naive Bayes method was used to determine the extent of examinees' ability on learned concepts[9]. Other studies classify the strength of examinees based on a bloom taxonomy. In another study, the population type was used as the basis for determining exam questions. In a homogeneous population and no information about the competency level of examinees, the problem was chosen with the difficulty level of the medium as a matter of beginning[10]. In a heterogeneous population and there is information about the grade level of the participants, select the problem with the level of difficulty of the medium according to the grade level. Rule-based methods are used to extract information on BPK RI audit report[11]. On previous research, rule-based methods combined with machine learning techniques it is proposed to assess the level of ability of regular students and students with special needs. Machine learning techniques used are Naive Bayes, Multilayer Perceptron, SMO, Decision Tree, JRIP, and J48. The best accuracy results are achieved using the K-Means rule-based method of 64.12 [12]. The rules for the determination of the level of ability are formed based on expert opinion. The strength of examinees to vary and the amount of data evolving lies in need for dynamic formation of rules[13]. The discovery of the pattern on the test data of the participants can be used as the basis for the establishment of states to replace the expert as well as improve the prediction accuracy. It is necessary to extract the pattern so that it can be used for the formation of the initial capability rules for examinees. This paper discusses the extracting ability of cluster-based exam participants with K-means for the determination of rules in the adaptive appraisal system[14]. The addition of K-means is aimed at exploring the participants 'group ability patterns from the participants' pretest answers and improving the accuracy of determining the ability level of examinees[15].

2. Methods

The ability level of examinees is determined based on preliminary test result data. Initial tests were performed to obtain test participants' ability data. Each test taker shows two types of tests, in writing and using math games. Problems given on the written test and experiments with the game are the same There are 150 records of test results and 150 records of test results using games. Each data record represents one test participant. Each record of written test results consists of 244 attributes. Attributes on the written test are the names, age, class, school, questions answered (60 items), correct answers (60 answers), competence related to elements (60 competencies), respondents' answers (60 responses) (60 status)[2]. Answer status is 0 or 1, with 0 for a wrong answer and 1 for the correct answer. The test results using the game consist of 304 attributes, 240 attributes equal to the write test attribute, plus the time attribute used to answer (60 times). The results of the written test were processed to gain the ability of the test participants and be ground truth after being validated by the math teacher. Initial test result data using the game is stored in the game log. Furthermore, the pattern extracting process is made by clustering process to find a profile of student's answer. The models that were seen were then analyzed for the establishment of rules for determining the level of competence of examinees[5]. Excavation pattern of student ability is made by using clustering. The clustering process is done to find the group pattern for each level. The K-Means method is used with the value k = 6. The selection of k values is based on the assumption that there are six grade groups according to grade level. The features used in the clustering process have six features. This feature states the value of each masked status examinees at grade 1, 2, 3, 4, 5 and six levels. This feature was chosen based on the results of previous studies that the most critical element in determining the ability level of examinees is the feature of graduation status of grades of each level. The results and cluster analysis formed are shown in Table I. The first pattern, 100000, means examinees have graduated degree 1 because they fulfill all competencies at level 1. Therefore, the participants can be categorized in the design of the ability of class 1. The second map, 111100, means the examinees have passed level 1, level 2, and level 3. This model can be categorized in the model of the ability of grade 3. The third model, 110100, meaning the test participants pass level 1 and level 2, so the examinees in this cluster can be grouped in a class 2 capability pattern.

Table I

Results And Classical Analysis Of Formed

Cluster	Middle value	Number of test takers	Persentase	Pattern class
Cluster 0	101000	63	45%	**1**
Cluster 1	111100	33	25%	**3**
Cluster 2	110100	19	13%	**2**
CLuster 3	101100	25	17%	**1**
Cluster 4	101000	5	5%	**0**
Cluster 5	**111100**	**6**	**5%**	**4**

The fourth pattern, 101100, means passing test participants graduate level 1 and level 3, but not graded scale 2, 4, 5, and 6. This model can be categorized as equivalent to the ability of class 1. In the next study, this model needs to be investigated further because there are 24 participants in it. The number of participants with this design cannot be considered as noise. The fifth model, 001000, means that examinees do not pass any level, so they can not be categorized in any class or can be called level 0. The sixth pattern, 111100, means the test participants pass level 1, 2, 3, and 4, so the participants in the group this can be categorized as having the ability equivalent to grade 4. Rules are formed from clustered class patterns has the following meaning. If the gradation pattern of the test participants is like the middle score of the fifth cluster (101000), then the test participants are categorized as having the ability equivalent to grade 0. If the gradation pattern of the test participants is like the middle grade of the first cluster (101000) or the fourth group (101100), then the test participants are categorized as having the ability equivalent to degree 1. If the graduation pattern of test participants is like the middle-value third cluster (110100), then the test participants are categorized as having the ability equivalent to class 2. If the gradation pattern of the test participants is like the middle score of the second cluster (111100), then the test participants are categorized as having the ability equivalent to grade 3. If the gradation pattern of the test participants is like the middle score of the sixth cluster (111100), then the test participants are categorized as having the ability equivalent to grade 4.

3. Result And Discuss

This section contains test results for the predicted ability level of test participants. Cross-validation and percentage split evaluation techniques were used to observe and analyze the results of predictive performance measurements using are Naive Bayes, Multilayer Perceptron, SMO, Decision Tree, JRIP, and J48. The testing steps are as follows. The test was conducted by applying six prediction methods using fold 10, 15, 20, 25, and 30 as well as 70%, 80%, 80%, and, 95% split percentage on the data processed using the rule-based method based on the expert.The test was conducted by applying six prediction methods using fold 10, 15, 20, 25, and 30 and 75% split percentage, 80%, 80%, and 95% on the data processed using the proposed method, an ie, rule-based approach based on clustering (termed Rule from Clustering (RC)). The analysis is done by comparing the prediction results on the numbers 1 and 2 using seven sizes classifier, i.e., Kappa, Mean Absolute Error (MAE), Precision, Recall, F-Measure, ROC, and accuracy. Table II and Table III show the predicted performance results using Naive Bayes (NB). Most performance measures optimal is to use folds 10, 15, 20, and 25 and use 95% percentage split on RC_NB. In folds 10, 15, 20, and 25 this Naive Bayes is able to reach the maximum value of Kappa = 0.97, MAE = 0.02, Precision = 0,75 Recall = 0.74, F-Measure = 0.75, and ROC = 0, 98.

Table II

Prediction Performance Measures Using Naive Bayes With Cross-Validation

Fold	Kappa	MAE	Precision	Recall	F-Measure	ROC
5	0,30	0,15	0,36	0,41	0,37	**0,74**
10	0,30	0,15	0,36	0,41	0,37	**0,72**
15	0,31	0,15	0,36	0,41	0,37	**0,72**
20	0,31	0,15	0,36	0,41	0,37	**0,72**
25	0,30	0,15	0,36	0,41	0,37	**0,71**
Metode RC_*Naive* Bayes (RC_NB)						
5	0,97	0,08	0,75	0,74	0,75	**0,98**
10	0,97	0,02	0,75	0,74	0,75	**0,98**
15	0,97	0,02	0,75	0,74	0,75	**0,97**
20	0,97	0,02	0,75	0,74	0,75	**0,98**
25	**0,97**	**0,02**	**0,75**	**0,74**	**0,75**	**0,98**

Table III

Prediction Performance Using Naive Bayes with percentage Split

Metode *Naive* Bayes (NB)						
PS	Kapp	MAE	Pr ec	Recall	F-Meas	ROC
75%	0,3	0,14	0,3	0,39	0,	**0,68**
80%	0,2	0,16	0,3	0,36	0,	**0,66**
80%	0,2	0,17	0,3	0,45	0,	**0,66**
95%	0,2	0,18	0,1	0,35	0,	**1,00**
Metode RC_*Naive* Bayes (RC_NB)						
75%	0,9	0,02	0,7	0,74	0,	**0,97**
80%	0,9	0,02	0,6	0,59	0,	**0,94**
80%	0,9	0,03	0,4	0,46	0,	**0,92**
95%	**1,0**	**0,10**	**0,5**	**0,51**	**0,**	**1,01**

While with 95% split percentage, Naive Bayes can achieve maximum value of Kappa = 1, 05 MAE = 0,10 and ROC = 1,01 although precision, recall, and F-Measure value are better when using 75% spent split compared to 95% split percentage. From these results, it can be concluded that the formation of rules based on clustering can improve the performance of predictions using Naive Bayes. The use of fold and percentage split alike shows a good improvement in yield compared to the rule-making method based on expert opinion. The predicted performance results using Multi-Layer Perceptron, are shown in Table IV, and Table V. The use of clustering to build proven rules can improve the performance of the Multi-Layer Perceptron method, as seen in the RC Multi-Layer Perceptron results. The use of fold 5, 10, 15, 20, and 25 on RC_Multi Layer Perceptron can improve performance with an average value of Kappa = 0.91 MAE = 0.02, Precision = 0.75, Recall = 0.74, F-Measure = 0, 76, and ROC = 0.98.

Table IV

Prediction performance measures using Multi-Layer Perceptron with cross-validation

Method MLP						
Fold	Kappa	MAE	Precision	Recall	F-Measur	ROC
5	0,35	0,15	0,40	0,42	0,41	**0,67**
10	0,33	0,16	0,36	0,33	0,33	**0,65**
15	0,34	0,15	0,39	0,42	0,40	**0,66**
20	0,39	0,15	0,45	0,47	0,45	**0,72**
25	0,30	0,15	0,35	0,44	0,38	**0,74**
Method MLP						
5	0,97	0,03	0,75	0,74	0,75	**0,99**
10	0,97	0,03	0,75	0,74	0,75	**0,98**
15	0,97	0,03	0,75	0,74	0,75	**0,98**
20	0,97	0,03	0,75	0,74	0,75	**0,99**
25	**0,97**	**0,03**	**0,75**	**0,74**	**0,75**	**0,98**

Table V

Prediction performance measures using Multi-Layer Perceptron with percentage Split

Method MLP						
PS	Kappa	MAE	Precision	Recall	F-Measure	ROC
70%	0,43	0,15	0,45	0,45	0,45	**0,78**
80%	0,30	0,16	0,22	0,26	0,24	**0,81**
80%	0,38	0,16	0,34	0,45	0,35	**0,68**
95%	0,29	0,18	0,36	0,39	0,37	**0,82**
Method MLP						
70%	0,95	0,02	0,62	0,58	0,59	**0,94**
80%	0,94	0,02	0,61	0,58	0,59	**0,94**
80%	0,92	0,02	0,48	0,45	0,46	**0,92**
95%	**1,00**	**0,01**	**0,50**	**0,50**	**0,50**	**1,00**

Performance improvements also occur in RC MultiLayer Perceptron with the use of percentage splits. As with RC_NB, the best Kappa and MAE values are achieved when 95% percentage split is used Kappa = 1, MAE = 0,01. However, Precision, Recall, and F-Measure values are best expressed when used 70% split percentage, respectively Precision = 0.62, Recall = 0.58, and F-Measure = 0.94. Table VI and Table VII show the predicted performance using SVM. With the use of fold 5, 10, 15, 20, and 25 on RC_SVM, the Kappa value increased to 0.96, Precision to 0.72, Recall to 0.74, F-Measure to 0.73, and ROC to 0.98, compared with SVM. However, the MAE score tends to be less competent, to 0.2, it is still better not to use clustering. The use of percentage split on RC_SVM also increases the value of Kappa, Precision, Recall, F-Measure, and ROC performance. Just like the use of fold, the MAE value tends to increase compared to the use of SVM for data processed by expert rules. The use of 75%, 80%, 85%, and 90% split percentage on the RC_SVM method is able to achieve the average value of Kappa = 0.96, Precision = 0.56, Recall = 0.59, F-Measure = 0.58, and ROC = 0.96. However, MAE performance measures decreased as MAE values rose to 0.23

Table VI
Prediction performance measures using SVM with cross Validation

Fold	Kappa	MAE	Precision	Recall	F-Measur	ROC
			Method SVM			
5	0,35	0,15	0,23	0,41	0,25	**0,69**
10	0,33	0,16	0,22	0,29	0,25	**0,68**
15	0,34	0,15	0,23	0,29	0,25	**0,69**
20	0,39	0,15	0,23	0,29	0,25	**0,68**
25	0,28	0,15	0,23	0,29	0,25	**0,67**
			Method RC_SVM			
5	0,96	0,21	0,72	0,74	0,73	**0,98**
10	0,96	0,21	0,72	0,74	0,73	**0,99**
15	0,96	0,21	0,72	0,74	0,73	**0,99**
20	0,96	0,21	0,72	0,74	0,73	**0,99**
25	**0,96**	**0,21**	**0,72**	**0,74**	**0,72**	**0,99**

Table VII
Prediction performance measures using SVM with percentage Split

PS	Kappa	MAE	Precision	Recall	F-Measure	ROC
			Method SVM			
75%	0,43	0,15	0,22	0,28	0,25	**0,63**
80%	0,30	0,16	0,21	0,28	0,24	**0,56**
85%	0,36	0,16	0,18	0,25	0,25	**0,59**
90%	0,29	0,18	0,13	0,21	0,16	**0,66**
			Metode RC_SVM			
75%	0,96	0,21	0,63	0,59	0,60	**0,95**
80%	0,96	0,21	0,62	0,59	0,58	**0,96**
85%	0,93	0,21	0,49	0,46	0,45	**0,96**
90%	**1,00**	**0,30**	**0,50**	**0,50**	**0,50**	**1,00**

4. CONCLUSION

Based on the results obtained, application of this method to student exam data proved to improve the performance of all the prediction methods used in this paper. Performance improvements are shown by an increase in Kappa, Precision, Recall, F-Measure, and ROC values, as well as a decrease in MAE values, except for the SVM method, the application of this technique tends to lead to an increase in cost MAE. This arrangement is suitable for adaptive appraisal system with the rules can be adjusted as the addition of the number of test data and the addition of the number of variations in the ability pattern of examinees. Also can be used to predict the level of ability of examinees, the results of this study can also be used to detect competencies that have been and have not been mastered by examinees. So even the participants who complete on average Cp6 also completed on Cp2. There is a possibility that there is an association relationship between competency and another competency. It can be observed and found with a significant amount of data and represents the entire class. This link is interesting to examine in subsequent research, by first adding datasets used for training and testing.

References

[1] A. Coates and A. Y. Ng, “Learning feature representations with K-means,” Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), vol. 7700 LECTU, pp. 561–580, 2012.

[2] G. Canas, T. Poggio, and L. Rosasco, “Learning Manifolds with K-Means and K-Flats,” Adv. Neural Inf. …, pp. 1–9, 2012.

[3] [M. Luz López García, R. García-Ródenas, and A. González Gómez, “K-means algorithms for functional data,” Neurocomputing, vol. 151, no. P1, pp. 231–245, 2015.

[4] T. M. Kantrowitz, C. R. Dawson, and M. S. Fetzer, “Computer Adaptive Testing (CAT): A Faster, Smarter, and More Secure Approach to Pre-Employment Testing,” J. Bus. Psychol., vol. 26, no. 2, pp. 227–232, 2011.

[5] M. J. Gierl and J. Zhou, “Computer Adaptive-Attribute Testing,” Zeitschrift für Psychol. / J. Psychol., vol. 216, no. 1, pp. 29–39, 2008.

[6] E. Guzmán and R. Conejo, “Self-assessment in a feasible, adaptive web-based testing system,” IEEE Trans. Educ., vol. 48, no. 4, pp. 688–695, 2005.

[7] J. Wang, J. Wang, J. Song, X. S. Xu, H. T. Shen, and S. Li, “Optimized Cartesian K-means,” IEEE Trans. Knowl. Data Eng., vol. 27, no. 1, pp. 180–192, 2015.

[8] Y. Li, K. Zhao, X. Chu, and J. Liu, “Speeding up k-Means algorithm by GPUs,” J. Comput. Syst. Sci., vol. 79, no. 2, pp. 216–229, 2013.

[9] Z. Huang, “Extensions to the k-means algorithm for clustering large data sets with categorical values,” Data Min. Knowl. Discov., vol. 2, no. 3, pp. 283–304, 1998.

[10] H. Nurdiyanto, R. Rahim, and N. Wulan, “Symmetric Stream Cipher using Triple Transposition Key Method and Base64 Algorithm for Security Improvement,” J. Phys. Conf. Ser., vol. 930, no. 1, p. 12005, Dec. 2017.

[11] R. C. De Amorim, “Constrained clustering with Minkowski Weighted K-Means,” in CINTI 2012 - 13th IEEE International Symposium on Computational Intelligence and Informatics, Proceedings, 2012, pp. 13–17.

[12] J. Ye, Z. Zhao, and M. Wu, “Discriminative K-means for Clustering,” Adv. Neural Inf. Process. Syst. 20, no. 1, pp. 1649–1656, 2008.

[13] D. Nofriansyah, “Application to Determination of Scholarship Worthiness Using Simple Multi-Attribute Rating Technique and Merkle Hellman Method,” Int. J. Artif. Intell. Res., vol. 1, no. 2, p. 41, Oct. 2017.

[14] A. P. Windarto, “Implementation of Data Mining on Rice Imports by Major Country of Origin Using Algorithm Using K-Means Clustering Method,” Int. J. Artif. Intell. Res., vol. 1, no. 2, p. 26, Oct. 2017.

[15] A. Sekmen and P. Challa, “Assessment of adaptive human-robot interactions,” Knowledge-Based Syst., vol. 42, pp. 49–59, 2013.