BEST
Journal of Applied Electrical & Science Technology – University of PGRI Adi Buana Surabaya
p-ISSN 2715-2871
e-ISSN 2714-5247

Implementation of Decision Tree and Item Response Theory (IRT) in SIMPATIF (Sistem Penilaian Adaptif)

Norma Devi Kurniasari
Informatics
Institut Teknologi Insan Cendekia Mandiri,
Sidoarjo Indonesia
normadevi@iticm.ac.id

Anis Shobikhah
Informatics
Institut Teknologi Insan Cendekia Mandiri,
Sidoarjo Indonesia
anisshobikhah@iticm.ac.id

Trian Basofi Rohman
Informatics
Institut Teknologi Insan Cendekia Mandiri,
Sidoarjo Indonesia
trian@iticm.ac.id

Muchamad Kurniawan
Informatics Engineering
Institut Teknologi Adhi Tama Surabaya
Surabaya, Indonesia
Muchamad.kurniawan.itats.ac.id

***Abstract*— This study discusses how to accurately determine the level of ability of students using Computer Adaptive Testing (CAT). In the implementation of CAT, test participants are given question items according to their abilities. This shows the difference from existing testing systems such as Computer Based Testing (CBT). ADAPTIVE in this system is an automated exam system that is carried out adaptively, adjusting the difficulty level of the questions to the ability of each examinee. The questions given depend on the answer to the previous question: true or false. If the answer is correct, then the next question item has a higher level of difficulty, while if it is wrong, the difficulty level of the next question decreases. CAT in this study is also called the term SIMPATIF (Adaptive Assessment System). The difference between CAT and SIMPATIF is from the selection of question items, in the test using the SIMPATIF application, the selection of question items using the C4.5 algorithm and IRT 3 Parameter Logistics Model (3PL). The 3 parameters are the difficulty level of the question, differentiation, and deceit. In the development of this SIMPATIF. The input attributes from the decision tree have 4 attributes, namely difficulty level, differentiation, deceiver and IRT, while the target attributes are 2, namely YES and NO. The questions used are productive subjects majoring in Multimedia at SMK PGRI 2 Sidoarjo. The study was conducted with 10 students, SIMPATIF only succeeded in 67,901% for the adaptive status of the questions. Meanwhile, if correlated with CBT 76.58%.**

Keywords*— component, Item Response Theory, Computer Adaptive Testing, Computer Based Testing, Paper And Pencil Test

I. INTRODUCTION

One way to determine the level of ability in a group in the field of education is by holding a test. The test can be done through various methods, including conventional methods and computer-based methods [1][2]. Tests are usually given in the form of multiple-choice questions, essays, projects, or presentations. Currently, conventional tests are starting to be abandoned, and computer-based tests are becoming increasingly sophisticated [3][4].

CBT (Computer-Based Test) is a computer-based testing system that was first implemented in Indonesia to replace the National Exam [5]. This transition has also affected the Final Semester Exam and Tryout tests conducted at SMK PGRI 2 Sidoarjo, which now utilize CBT. However, in its implementation, CBT has a disadvantage — it is less able to accurately determine students' levels of ability [6][1].

To address this limitation, the SIMPATIF application was developed as an alternative to CBT. SIMPATIF aims to determine the ability level of test participants using the Item Response Theory (IRT) with three logistic parameters (3PL): the difficulty level of the questions, the ability to discriminate between levels of ability, and the guessing parameter [1][5]. The core of this theory is the probability that a participant answers an item correctly, based on their ability level (θ) calculated beforehand.

Each item answered involves specific calculations under the IRT model. Furthermore, SIMPATIF incorporates the C4.5 decision tree algorithm to enhance the accuracy and classification of participant abilities [7][8][9]. The SIMPATIF application was developed as a web-based platform, operating online. The question items used in the system's question bank were sourced from historical data of previous multimedia productive exam questions [10][11][12][13].

.

II. METHODS

The estimation of test takers' ability in CAT is carried out based on IRT. It is assumed that at the start of the test, test participants are given questions with a moderate initial difficulty level (initial b = 0). If the questions with a moderate level of difficulty can be answered correctly, participants are given new questions that are more.

$$\theta = b_i + \frac{1}{Da_i}\ln(0.5(1 + \sqrt{(1 + 8c_i)} \quad \text{.....(1)}$$

Information:

θ: test taker ability level

ai : Distinguishing Power Index Item I

Bi : Difficulty Index Item I

Ci : Index of Pseudo-Guessing Item I

D : the scaling factor which costs 1.7.

If θ is already known, then Pi (θ) or the probability of the test participant's ability to answer correctly can be calculated. If Pi (θ) has been calculated, then the probability of answering incorrect Qi (θ) can be calculated

p-ISSN 2715-2871
e-ISSN 2714-5247

with the formula:

$Qi(\theta) = 1 - Pi(\theta)$..... (2)

If the probability of answering correctly Pi (θ) and the probability of answering wrong Qi (θ) are known, then the standard error of the measurement SE(θ) can be calculated using equation 3.

$$SE(\theta) = \frac{1}{\sqrt{\sum_{i=1}^{N} I_i(\theta)}} \quad \text{..... (3)}$$

If the process of presenting question items in CAT is carried out repeatedly with an index of difficulty of items that are in accordance with the ability of the test participant, then in the end it can be obtained that the standard error of the measurement SE(θ) is increasingly constant and the absolute price of the difference in SE(θ) between the repetitions of the question presentation is getting smaller. In equation 3 there is Question Item Information with symbol (Ii (θ)) which can be calculated with equation 4.

$I_i(\theta) = Pi(\theta)Qi(\theta)$ ..... (4)

If the absolute value of the difference SE(θ) between the repetitions of the question presentation has reached the smallest limit determined (<= 0.01), then the process of presenting questions to test takers in CAT can be stopped. The way to stop presenting this question is called the stopping rule or stopping criterion.

In principle, the stopping rule or stopping criterion ensures that the number of questions presented in the CAT must be limited. The limitation on the number of questions can be because the question items are exhausted, the length of the test has been achieved, the level of accuracy of the ability estimate has been Final estimate of the ability of the test participant. Stopping rule has many alternatives according to your needs system, in this study is combined with the C4.5 algorithm, so the addition of stopping rules apart from the CAT algorithm itself also occurs when the presentation of question items does not have a pattern according to the rule.

The ability of the test participant (θ) and also the difficulty level of question item (b) in IRT can be positive or negative. In this study, the lowest ability of test participants (θ) was limited to -3 and the highest was +3. The reason for this restriction is that under normal conditions or distribution data smaller than -3 or greater than +3 is very small.

In general, the c4.5 algorithm for building a decision tree is as follows:

1. Select the attribute as the root.
2. Create a branch for each value.
3. Divide cases within branches.
4. Repeat the process for each branch until all cases on the branch have the same class.

An example of a decision tree with a SIMPATIF dataset is like figure 1 below.

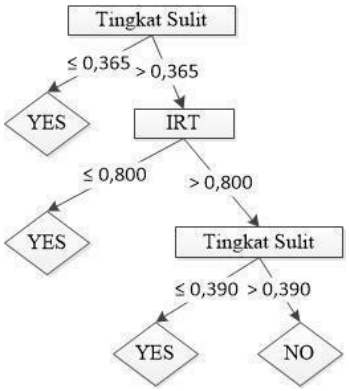

Fig. 1. Dataset Decision Tree Of SIMPATIF

The decision tree is processed from the Rapidminer datamining software. Then it is implemented in the SIMPATIF application. The decision tree in the form of a rule can later be replaced according to the needs of the test participants. In the SIMPATIF application, there are 3 access roles, namely admin, teacher and student. The system architecture in Figure 2 focuses more on the flow of the system during the execution of the exam. The image shows that admins have an important role in preparing data before it is used as knowledge. The students' answer data is preprocessed first and then the training process is carried out using data mining software, namely Rapidminer. Rapidminer is a software Classification that is open source for the field of education. The advantages of this software are a more varied use and a more user-friendly interface

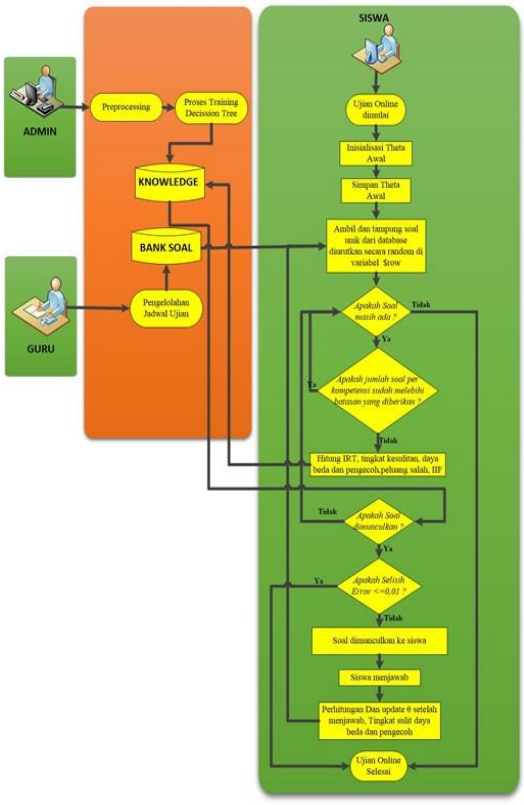

Fig. 2. System Architecture

From figure 2, it is divided into 3 stages, namely the first preprocessing, the second is the training process with

p-ISSN 2715-2871
e-ISSN 2714-5247

Decision Tree, and the third is the Online Exam itself whose application is called SIMPATIF.

Processing is discussed in the next subchapter, after this stage, a training process is carried out with the Decision Tree C4.5 classification method using Rapidminer software. It is at this stage of the online exam that the CAT algorithm and decision tree are used, the first step for the teacher to manage the online exam schedule according to the subject and class. The questions are taken from question banks that already have the parameters of the difficulty level, tricksters, and differentiation. If the question does not have these 3 parameters, then the question cannot be included in the question bank. The next step, test takers or students conduct an online exam, in figure 2 for this testing stage there are several stages that will be explained in the following subchapter.

1. Starting Rule (Procedure for selecting the first question item)

The first stage is the initialization of the ability level (θ) of the initial test participant by being given an ability of 0.0 which means that the initial ability is medium. The value of the ability is stored for the calculation of IRT or the probability of the participant answering correctly. The system will take question items from the question bank and store and sort them randomly on $row variables. Competency Standards (SK) are also randomized. means that test takers will do randomly for each SK.

2. Procedure for selecting question items during the test using the rule of the C4.5 algorithm

Before the question item is presented to the test participant, the system first checks whether there are question items in the question bank, if there are no question items then the exam is dismissed, if there are still then the system will check whether the number of question items per competency has exceeded the limit that has been given, if it has exceeded the limit then repetition occurs when checking whether there are questions or not. If it has not been exceeded, then calculate the IRT, difficulty level, differentiation, deception, wrong chance and IIF. Then the values are sent to the C4.5 rule decision tree to be analyzed whether they appear or not. So the system checks whether the question item is raised or not, if not, it will return to the stage of checking whether there are still questions or not.

3 Procedure for Stopping Tests

In this SIMPATIF application, the test will end if the SE difference condition is <=0.01. If the error difference is still not reached, the question will be presented to the test participant. The second is when the 3 parameters of the question item and the value of IRT or Pi are not detected in the rule or knowledge. Third, if the number of questions has reached the limit. The fourth test will stop if the specified time has expired. When a question is raised, the test taker must answer the question item, because each question item will be taken into account for its ability level, the test taker cannot repeat the previous question item and cannot see the next question item.

After the test taker answers the question items, the system will update the ability value (θ) after answering, the level of difficulty, differentiation and deception. Then the system will repeat the first step, namely the system takes question items from the question bank and accommodates and sorts them randomly on $row variable. If you have met the error then the exam will be dismissed.

3. Estimating Students' Abilities

The estimated ability of students in the SIMPATIF application is calculated on each question item worked on by the test participants. The value of this ability is calculated along with updating the difficulty, differentiation and deception levels

III. RESULT AND DISCUSSION

The research was held at SMK PGRI 2 Sidoarjo, with 10 students. The 10 students were given an online exam using CBT and SIMPATIF. With the criteria of CBT as many as 40 questions and SIMPATIF 40 questions. The following table is a table of CBT and SIMPATIF results

TABLE 1 RECAP OF ADAPTIVE SCORES AND STATUS

No.	NIS	Value	Number of Questions	Adaptive Score
1	11001	65	23	60.86
2	11002	86	21	66.66
3	11004	76	17	64.70
4	11008	82	22	72.77
5	11011	95	21	76.19
6	11016	87	19	65.21
7	11019	83	23	65.21
8	11027	86	20	63.63
9	11029	93	14	78.57
10	11036	52	23	65.21

In students who have high scores, namely students 1011 and 11029, they also have high adaptive scores, namely 76.19 and 78.57. For low scores, namely 11036 students, have a low adaptive score of 65.21%. It can be concluded that the higher the student's score, the more the adaptive score increases.

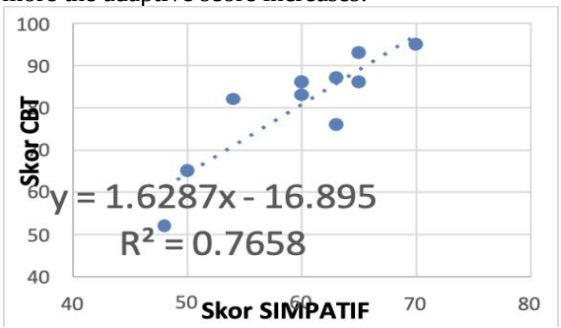

Fig. 3. Correlation – Comparison of CBT vs SIMPATIF

From figure 3, the results are unsatisfactory, the linear trend line moves upwards so that it can be said that the value is directly proportional. The value of the Coefficient of Determination (R2) is 0.7658 indicating

p-ISSN 2715-2871
e-ISSN 2714-5247

76.58 % (0.7658*100) the existence of the relationship between the y axis (SIMPATIF) and the x axis (Pure CAT) is still not good, the best value is at least 80%.

IV. CONCLUSION

Based on the thesis development process, which involved stages from research to trials, several conclusions were obtained. The SIMPATIF system showed a correlation of 76.58% with the CBT system, indicating a fairly high level of association. SIMPATIF also successfully met the research objective, which was to be able to select question items that would appear in the exam with an accuracy rate of 67.901%. When the standard error (SE) difference calculation was removed, all questions appeared, and the assessment criteria based on decision tree rules were not applied. The main advantage of SIMPATIF lies in its efficient processing time, as students do not need to answer all exam questions. In addition, the results obtained by learners are stored and reused as parameters to determine the format of future tests. For future development of SIMPATIF to make it better and more efficient, it is recommended to conduct a more in-depth analysis of the question difficulty levels, as these significantly affect the appearance of questions during exams. It is also necessary to test the attributes of the decision tree so that it can generate optimal hypothesis values.

REFERENCES

[1] Suwarto, 2011, *Classical Test Theory and Modern Test Theory*
[2] Arikunto Suharsimi, 2008, Basics of Educational Evaluation (Revised Edition), Bumi Aksara
[3] Erhan Delen, 2015, Enhancing a Computer-Based Testing Environment with Optimum Item Response Time, Eurasia Journal of Mathematics, Science & Technology Education
[4] Crisp Geoffrey, 2007, The e-Assessment Handbook, Continuum International Publishing Group
[5] Djemari Mardapi, Haryanto, Samsul Hadi, 2012, Development of Computerized Adaptive Testing System, LPPM UNY
[6] Crocker, L. and Algina, J., 1986, Introduction to Classical and Modern Test Theory, New York: Holt, Rinehart and Winston, Inc.
[7] Wahyudin, 2009, Iterative Dichotomizer 3 (ID3) Method Mining, Yogyakarta: CV. ANDI OFFSET
[8] Susanto, S., & Suryadi, D., 2010, Introduction to Data
[9] Goodness, Santosa, 2007, Data Mining Techniques for Data Utilization for Business Purposes, Graha Ilmu. Yogyakarta
[10] Del Sorbo Maria Rosaria & Walter Balzano, e-Xamina: an Experimental Multi-user Assessment Platform for Computer Adaptive Testing, IEEE
[11] Phankokkruad Manop, 2012, Association Rules for Data Mining in Item Classification Algorithm: Web Service Approach, IEEE
[12] Junhua Chen, Wei Lai & Tang Ying, 2013, A Secure and High Efficient Computerized Adaptive Testing Scheme for Computer Courses Based on Identity, IEEE
[13] Ueno Maomi & Pokpong Songmuang, 2010, Computerized Adaptive Testing Based on Decision Tree, IEEE