Advances in Social Science, Education and Humanities Research, volume 474
Proceedings of the 2nd International Seminar on Science and Technology (ISSTEC 2019)

Improving the Accuracy of Student Problem Identification Using *Rule-Based Machine Learning*

Budi Sulistiyo[1,*], Bayu Surarso[2], Wahyul Amien Syafei[3]

[1]*Graduate School of Information System Diponegoro University Semarang, Indonesia*
[2]*Faculty of Science and Mathematics Diponegoro University Semarang, Indonesia*
[3]*Department of Electrical Engineering Diponegoro University Semarang, Indonesia*
[*] *Corresponding author: budimantaro@gmail.com*

ABSTRACT
Adolescence is a period of development that is prone to problems and often makes adolescents unable to control emotions. No exception to adolescents who are in high school education. Problems that do not need to be resolved immediately and will arise even greater problems later on. Many methods of solving students' problems are carried out in conventional ways that require relatively takes time and costly. Therefore teacher career guidance and policy in schools use the problem list method provided for students. One thing that promises to improve accuracy with a short time to identify students' problems by creating information systems using intelligent technology such as machine learning. Machine learning offers sophisticated techniques in creating automated schemes that can be used by students and guidance counseling teachers in technical issues is on the rise. This article discusses issues related to learners but also offers knowledge-based users (rules) that can be used by counseling guidance teachers to replace those who are behind information systems. The results of this study indicate that the information system developed which is based on rule-based machine learning offers a classification that is more accurate, faster, can be done anytime, anywhere and requires no cost compared to existing conventional methods.
Keywords: *student problems, information systems, rule-based classifier, machine learning.*

1. INTRODUCTION

Problems experienced by humans will be increasingly complex with age, especially in adolescents who are still in high school. During this high school, students begin to recognize the wider environment that is faced with more difficult problems that require more serious handling. Problems experienced by students need to be addressed immediately because they are related to achievement [1]. To prevent greater problems arising in the future, an effort is needed to detect students' problems from the start. Benefits in understanding the involvement of behavior problems among youth and informing the design of interventions to reduce behavior problems [2].

There are several ways you can do techniques in gathering data about student behavior problems such as interviews, observations, questionnaires, and problem checklists. The method used to identify students' problems is done in a conventional manner which is sometimes still subjective, must meet with the counseling teacher, done in class, requires a relatively long time and cost. The use of problem checklists to identify student problems is done in a conventional way that requires a relatively long time and cost. The checklist uses indicators that can be a sign of the degree of behavior that can be measured. These indicators are represented by items given in the questionnaire format for later classification. The classification process depends on the rules made by humans with the appraisal function adding the score of related items in the questionnaire in calculating the results. Therefore the quality of the results of the classification is determined in (1) Knowledge and user experience used in the classification process (2) Items used in this technique (3) Rules in the valuation function [3].

In calculating the score in the questionnaire, the component needs to be designed rules that require the experience and knowledge of an expert in their field. Change the rules that have been made from the knowledge of previous cases to improve accuracy in the identification process. By learning from data and past experience to optimize results such as machine learning [4]. By using many examples of data and machine learning rules can be used to predict data that is not yet clear [5].

Machine learning is an algorithm that can learn from data and make predictions without explicit deprogrammed to do something [6]. With the help of developing intelligent systems that are able to make decisions independently. Machine learning algorithm learns from past data examples through pattern matching, then based on the data learned will predict the results obtained. In the process of identifying the problem of students include whether the individual is in trouble or not by using items that have been determined and also classification, in this case, can be done with the supervised machine learning method.

The results of this study indicate that the information system developed which is based on rule-based machine learning offers a classification that is more accurate, faster, can be done anytime, anywhere and requires no cost compared to existing conventional methods.

2. MATERIALS AND METHODS

2.1 Materials

For making rule-based machine learning information systems the material used in this study is questionnaire data from the student's problem checklist (DCM) that is used by counseling guidance teachers in identifying problems that have or are being experienced by students. Data collection uses measurement tools, in this case, related to psychology using the inventory method. Inventory is a type of survey or questionnaire used to measure respondents. As measured as individual personal circumstances such as attitudes, interests, physical conditions, psychiatric social relationships, conditions of the home and family and so forth.

The tool used in this study is a laptop with the following specifications:

1. Lenovo Ideapad 300 Intel® Core ™ i7-6500U Laptop CPU @ 2.50GHz (4CPUs), ~ 2.6GHz
2. Memory 16384MB
3. 240GB hard drive
4. Display Intel® HD Graphics 520
5. Windows 10 Pro 64-bit (10.0, Build 18362)

The programming language used in this study:

1. Node.js.

Node.js is software designed to develop web-based applications written in the javascript programming syntax. Javascript usually runs on the client or browser side only, but because it has been found, javascript node.js can be used as a programming language on servers such as PHP, Ruby ASP, C #, and others. In other words, node.js is needed to be able to execute javascript outside the browser, so that with node.js the developer can manage the client and server sides themselves [7].

2. Express.js.

Express.js is a framework used to facilitate the creation of web-based applications and mobile applications. The minimal and flexible express.js framework provides a powerful set of features that can be used to create web applications more easily and quickly.

3. XAMPP.

XAMPP is computer software that acts as a server consisting of several programs including Apache HTTP Server, MySQL database, and language translators written in the PHP and Perl programming languages.

4. MySQL.

MySQL is a multithread database management system and multi-user SQL or DBMS that uses SQL to access it. SQL can be interpreted as a language used to access data in a relational and structured database, while MySQL, in this case, becomes a software or tool for managing or managing SQL by using Queries.

5. Visual Studio Code.

Visual Studio Code is a text editor developed by Microsoft that is multiplatform, which can be used in a variety of operating systems. Visual Studio Code has thousands of extensions that can be accessed directly in Marketplace-VSCode which can help in the process of making programs easier.

2.2. Method

2.2.1. Rule-Based Machine Learning

Rule-Based Machine Learning (RBML) in terms of computer science that shows machine learning that is used for a particular purpose by using "rules" with existing data [8]. This rule has a general format consisting of if-then rules that make it easy for users. The method offered in identifying students' problems based on the architecture is shown in Fig. 1. The RBML method uses rules in the search. RBML evaluates the rules found and removes those that are not used so that it will reduce the data search space.

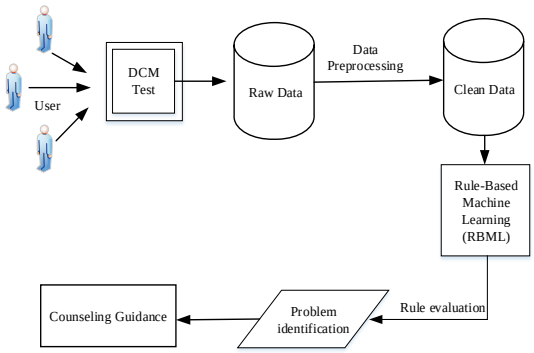

Figure 1. The proposed RBML architecture in identifying students' problems.

Data is collected by an application called the DCM Test that implements a number of questions in 12 different problem categories (Health, Economic Situation, Family Life, Religion and Moral, Recreation and Hobbies, Personal Relations, Social Life Organizations, Activities, Youth Problems, School Adjustments, Adjustments Curriculum, Study Habits, Future and Life Services Education). After the raw data is obtained, pre-processing operations are carried out, eliminating data that is not used. After clearing data from unneeded data, a learning algorithm is applied to find a set of rules that represent the relationship between the variables in the training data set and the class variables. The data set is then evaluated according to the rules in the training instance. The evaluation results are a classification system that will be used to identify class grades for cases that have not yet been classified. The results of the

identification of this problem can be used by students to get guidance and counseling for further processing.

2.2.2. Synthesis

This research is divided into 2 stages: the first stage is the beginning of the study. At this stage, the first step is the study of literature based on research that has been carried out such as: literature on machine-based learning Rule (RBML) is part of the field of computer science that extracts "rules" in solving problems using data [8]. Literature about rule-based classification, Rules are expressed in the form of if..then statements, rules are generally stated in the form "IF conditions conclusions, in this case, IF referred to as" antecedents "or" conditions "and THEN is referred to as" consequences "or" conclusions "[9]. In rule-based learning, the process partition used focuses on identifying the subgroup models contained in the training set [10]. The aspect that determines RBML in data mining is how to identify a set of rules that represent knowledge in the data to make decisions. Using machine learning can improve decision making; it has improved better results in time efficiency and predictive accuracy [11]. Literature about social, emotional and behavior problems that affect the lives of high school students. Literature lessons are conducted to reinforce the theories used in research The second part of Field studies, at this stage, use the step of collecting data directly into the field with observations for locations. then, discuss and interview students, counseling teachers or psychologists, vice-principals in the curriculum section and vice-principals in the student section.

The second stage is the stage of developing this research system using the SDLC waterfall model which is a software development process through phases such as a waterfall (flowing down) that must be carried out in order to succeed [12], shown in Figure 2.

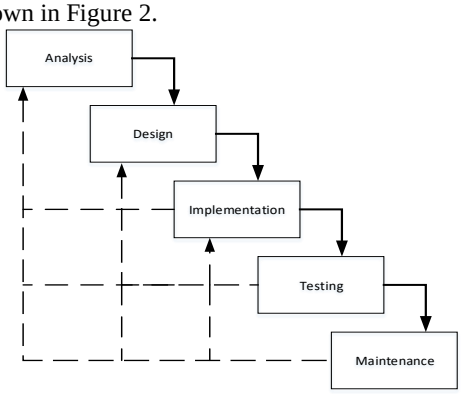

Figure 2. The Waterfall Model.[12]

1. Analysis

At this stage to determine the information system requirements as expected user use and system restrictions. Information can be obtained by interview or discussion, survey and then documented for use at a later stage.

2. Design

This stage is carried out system design with modeling tools, data structures, database design, and display system design.

3. Implementation

This stage is done by making coding using programming software. Making an information system is broken down into smaller modules which will later be combined again at a later stage. Each module is checked, whether it has fulfilled the desired function or not.

4. Testing

This stage is done by combining all modules that have been made and then testing the software if it is in accordance with the design and there are errors (bugs). If there are differences or errors with the design, it must be adjusted to the design and must be corrected.

5. Maintenance

This is the last stage in making information systems. The finished system is run down and needs to be maintained. Maintenance is included in error repair if an error is found in the previous steps.

3. RESULTS AND DISCUSSION

Data is collected by information systems from problem checking data (DCM). After the raw data is obtained, the data will be processed first to process the selected data. After the data is cleared of data that is not used, the learning algorithm is applied to find a set of rules that connect data set variables and class variables. The data set is then evaluated and grouped into classes that have been stored in the training rules. Rule-based machine learning will find the ratio of data sets to the number of problem items, the results of the ratio are then evaluated with the rules that have been saved. The result of rule evaluation is a classification information system that will be used to identify students' problems as shown in Table 1.

Table 1. Scale Problem.

Skala Persentase %	Value	Description
0	A	Very Good
1 - 10	B	Good
11 - 25	C	Fair
26 – 50	D	Poor
51 - 100	E	Very Poor

When users answer questions that are displayed, users are offered to answer between yes or no. Class labels will be automatically assigned during data collection by the final score method obtained after screening. There are two values from the class, namely "1" if the user experiences a problem and "0" if there are no problems asked by the system. The results of the answers will then be classified by the class during training. The results from the scoring of each category will then be calculated based on the number of questionnaires and then the percentage. From the results of the presentation, it will be on the scale of the rules, and the values will be generated from each class as in Table 2.

Table 2. Scoring Problem Category.

PROBLEM CATEGORY		TYPES OF PROBLEMS PROBLEM ID							AMOUNT	%	Value
I	**PERSONAL**								12	12,00	C
A	HEALTH	3	11	12	15	0	0	0	4	20,00	C
B	ECONOMIC SITUATION	28	0	0	0	0	0	0	1	5,00	B
C	FAMILY LIFE	43	46	0	0	0	0	0	2	10,00	B
D	RELIGION AND MORAL	66	0	0	0	0	0	0	1	5,00	B
E	RECREATION AND HOBBY (STATE)	81	85	89	91	0	0	0	4	20,00	C
II	**SOCIAL**								9	15,00	C
A	PERSONAL RELATIONSHIP	109	118	119	120	0	0	0	4	20,00	C
B	SOCIAL LIFE - ORGANIZATIONAL ACTIVITY	128	133	136	0	0	0	0	3	15,00	C
C	YOUNG PROBLEMS	153	158	0	0	0	0	0	2	10,00	B
III	**STUDY**								7	11,67	C
A	ADJUSTMENT OF SCHOOL	177	179	0	0	0	0	0	2	10,00	B
B	ADJUSTMENT OF CURRICULUM	194	0	0	0	0	0	0	1	5,00	B
C	LEARNING HABITS	201	202	210	218	0	0	0	4	20,00	C
IV	**CAREER**								1	5,00	B
A	FUTURE AND EDUCATION / SERVICE LIFE	223	0	0	0	0	0	0	1	5,00	B

Figure 3. Shows the results of performance in the identification of student problems obtained by applying rule-based classification in 12 categories. By looking at the scale of the problem, it can be identified the problems being faced by students in Health, Recreation and Hobbies (State), Personal Relationship, Social Life - Organizational, Learning Habits. With the problems that have been identified, students can be given priority guidance or counseling in the social field first, personal, study and then career, as shown in Figure 4.

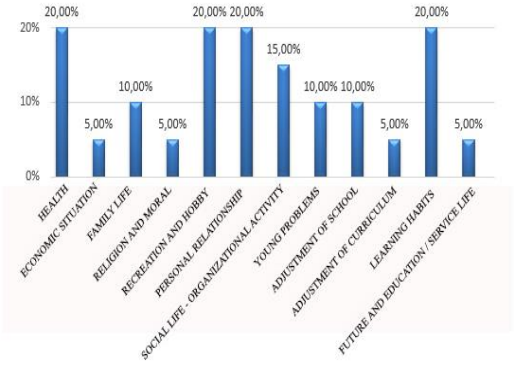

Figure 3. Problem Topic.

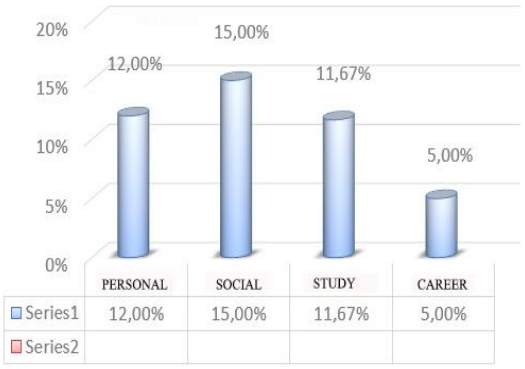

Figure 4. Guidance type.

4. CONCLUSION

Adolescence is a period of growth and development that experiences increasingly complex problems, especially when they are at the age of high school education where they begin to get to know the environment and the wider community. If a problem that is being faced by a student is not resolved immediately, it will cause an ongoing problem, and will also affect student achievement. In understanding and identifying problems experienced by participants is the first step in handling further. The methods used in understanding students such as questionnaires, DCM and others. The assessment is done using the calculated answer score. During classification and assessment still use conventional methods which are sometimes still subjective. So that one of the important problems in research identifying students' problems is to improve the classification process so that the services in guidance and counseling are more accurate and faster. To achieve this, a rule-based machine-based information system is used that builds an accurate classification of data and previous cases. This article proposes a rule-based machine learning information system that is not only accurate but also fasts in the classification process in automatically identifying student problems that can be exploited by students and teachers.

ACKNOWLEDGMENT

This work was supported by DRPM, Ministry of Research, Technology, and Higher Education of the Republic of Indonesia through the LPPM Diponegoro University under the Magister Thesis Research schema Year 2019 (Grant Number 258-18/UN7.P4.3/PP /2019).

REFERENCES

[1] McLeod, J. D., Uemura, R., Rohrman, S. (2012). Adolescent Mental Health, Behavior Problems, and Academic Achievement. Journal of Health and Social Behavior. 53(4): 482-497.

[2] El-Shenawy, O. E., Shehata, A. M. (2014). Applying Problem Behavior Theory in a Developing Arabic Country: Egypt. SAGE Open. 4(1): 2158244014521819.

[3] Thabtah, F., Peebles, D. (2020). A New Machine Learning Model Based on Induction of Rules for Autism Detection. Health Informatics Journal. 26(1): 264-286.

[4] Baştanlar, Y., Özuysal, M. (2014). Introduction to Machine Learning. In Introduction to Machine Learning. London. England. 1107: 105–128.

[5] Sudharsan, B., Peeples, M., Shomali, M. (2014). Hypoglycemia Prediction using Machine Learning Models for Patients with Type 2 Diabetes. Journal of Diabetes Science and Technology. 9(1): 86-90.

[6] Bleidorn, W., Hopwood, C. J. (2019). Using Machine Learning to Advance Personality Assessment and Theory. Personality and Social Psychology Review. 23(2): 190-203.

[7] Shah, H., Soomro, T. R. (2017). Node. Js Challenges in Implementation. Glob. Journals Inc. 17(2).

[8] Lungaroni, M., Murari, A., Peluso, E., Vega, J., Farias, G., Gelfusa, M., Contributors, J. E. T. (2018). On the Potential of Ruled-Based Machine Learning for Disruption Prediction on JET. Fusion Engineering and Design. 130: 62-68.

[9] Aggarwal, C. C. (2014). Data Classification Algorithms and Applications. New York. 2–6.

[10] Bassel, G. W., Glaab, E., Marquez, J., Holdsworth, M. J., Bacardit, J. (2011). Functional Network Construction in Arabidopsis Using Rule-Based Machine Learning on Large-Scale Data Sets. Plant Cell. 23(9): 3101–3116.

[11] Liu, Y., Zhao, T., Ju, W., Shi, S. (2017). Materials Discovery and Design using Machine Learning. J. Mater. 3(3): 159–177.

[12] Bassil, Y. (2012). A Simulation Model for the Waterfall Software Development Life Cycle. 2(5).