Student Behavior Analysis to Detect Learning Styles in Moodle Learning Management System

1st Yunia Ikawati
Departemen Teknik Informatika dan Komputer
Politeknik Elektronika Negeri Surabaya
Surabaya, Indonesia
yuniaa.ikawati@gmail.com

2nd M. Udin Harun Al Rasyid
Departemen Teknik Informatika dan Komputer
Politeknik Elektronika Negeri Surabaya
Surabaya, Indonesia
udinharun@pens.ac.id

3rd Idris Winarno
Departemen Teknik Informatika dan Komputer
Politeknik Elektronika Negeri Surabaya
Surabaya, Indonesia
idris@pens.ac.id

***Abstract*— E-learning is distance learning that uses computer technology, networks of computers and the internet. E-Learning allows students to study via computers in their respective places without having to go to study/lectures in class physically. Moodle is a Learning Management System that is used as a medium for delivering E-Learning. The problem that often arises in e-learning is that in the learning process, students interact more with e-learning media so that teachers will find it difficult to monitor student behavior when using learning media. In fact, students in some cases tend to drop out or attend lesser classes. Moodle can capture student interactions and activities while studying online using log files. From the results of student interactions and activities on e-learning, it can be used to determine their learning style. Identifying student learning styles can improve the performance of the learning process. This research suggests an approach to automatically predicting learning styles based on the Felder and Silverman learning style (FSLSM) model using the Decision Tree algorithm and the ensemble Gradient Boosted Tree method. We've used actual data sets derived from e-learning program log files to perform our work. We use precision and accuracy to assess the results. The results show that our approach is delivering excellent results**.

Keywords*— *E-Learning, Moodle, Learning Management System, Learning Style, Felder and Silverman learning style model, Decision Tree, Ensemble Method, Gradient Boosted Tree.

I. Introduction

The growth of internet science and technology in some fields is getting quicker. This advancement has shifted the nature of society in terms of seeking and receiving information. It is no longer confined to media, audiovisual and electronic content, and other sources of knowledge, one of which is through the internet network. The fast growth of this technology also affects on World of learning. Where learning is a process of communication and information to students from educators. This requires education information, initially on education. The process was carried out directly from face to face, but now An education system called e-learning has been developed [1].

E-learning is distance learning that uses computer technology, computer networks, and the internet. E-Learning allows learners to study through computers at their respective locations without attending classes/lectures in the classroom physically. E-learning is often also interpreted as a type of web-based learning that can be accessed through intranets on local networks or the internet. One of the technologies that are present to develop an E-Learning system is the Learning Management System (LMS) [1]. Learning Management System (LMS) is a form of online learning presentation (e-learning). LMS generally have features similar to conventional learning such as providing teaching materials, discussion forums, quizzes and assessments. Some LMS that are widely used includes Moodle [1]. Moodle's use has been cited by many works of literature as a valuable method for education, developing student study and critical thinking skills, facilitating independent study, and supporting group practices. [2]. LMS is also used to monitor student attendance activities, record a user's login time as a marker of attendance at E-Learning and provide information about user activity when using the E-Learning application [3]. The problem that often arises in E-Learning learning is that students interact more with E-learning media in the learning process, so teachers will have difficulty monitoring the behavior of students when using learning media. Even in some cases, students tend to drop out of class or lack of class attendance. Also, active learning methods for each person are different, so educators must see what the "learning preferences" of students are. So it is expected that knowing what the majority's preferences will facilitate educators in presenting the material [4].

From the previous problem, it is necessary to analyze student behavior (Behavior Analysis) on the E-learning system to find out appropriate learning styles. Learning style refers to the chosen way students learn, process, interpret, and store knowledge [5]. In this study, we developed a new analysis to automatically detect student learning styles based on the Felder Silverman Learning Style Model by analyzing student behavior towards E-Learning using the decision tree method and ensemble method Gradient Boosted Tree.

Several studies related to the student learning styles detection have been conducted. Hasibuan et al. [4] detect learning styles based on prior knowledge from students using the Neural Network algorithm referring to the VARK learning style. Hoang et al. [5] identify learning styles based on questionnaires using neural network methods and statistics relating to FLSM learning styles. Hasibuan et al. [6] automatically detect learning styles based on student log data using a hybrid method referring to the VARK learning style. Wang et al. [7] propose an automatic method for identifying learning styles based on the K-Means algorithm, then an

978-1-7281-9530-8/20/$31.00 ©2020 IEEE

analysis of E-Learning recommendations based on the Page Rank algorithm and correlation analysis of the learning effects referring to the RFL learning style. Quafae et al. [8] [9] automatically detects learning styles based on log data from students using the K-Means and Naive Bayes methods referring to FSLSM learning styles. Manal et al. [10] identify learning styles from log data analysis using decision trees referring to FSLSM learning styles. Ling et al. [11] detected learning styles from log data analysis using augmented naive Bayes tree referring to FSLSM learning styles. Umi et al. [12] automatically detect learning styles based on log data from students using the SOM method.

This research proposes a new approach to detect students' learning styles dynamically based on their behavior in e-learning. The results will be compared to a questionnaire based on the Learning Style Model of Felder Silverman, using the decision tree method and the Gradient Boosted Tree method of the ensemble. The experimental results show that the proposed method is efficient.

II. System Design

General description of the research design in this study can be seen in the following Fig. 1:

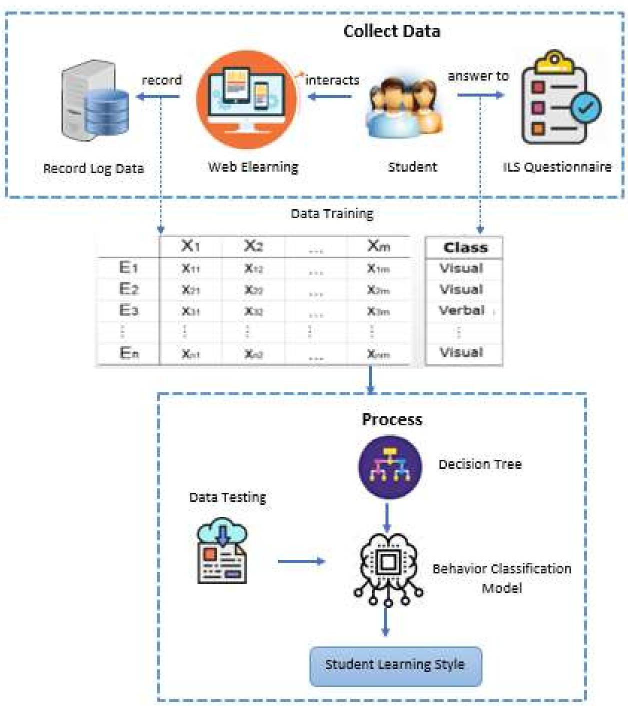

Fig. 1. System Design

Fig. 1 shows the research system design. This research consists of collected data and process. Collect Data is a process to collect student behavior from learning obtained from moodle log data. All Moodle derived information is collected from the Moodle data log of the course [13]. So the results from the moodle log data will be selected and used as an attribute of the dataset. After we label the FSLSM-based education resources, each attribute is the sum of the students' total visits (and other information) to each learning object. Students also fill in the ILS questionnaire. Each object receives a label that matches the student's Learning Style - identified through the ILS questionnaire. The results of the dataset are used as training data. The process component involves methods and techniques for processing training data using different methods, in this study using the approach of classifying it using the decision tree. The result of this process is the detection of students' learning styles based on the behavior of E-Learning.

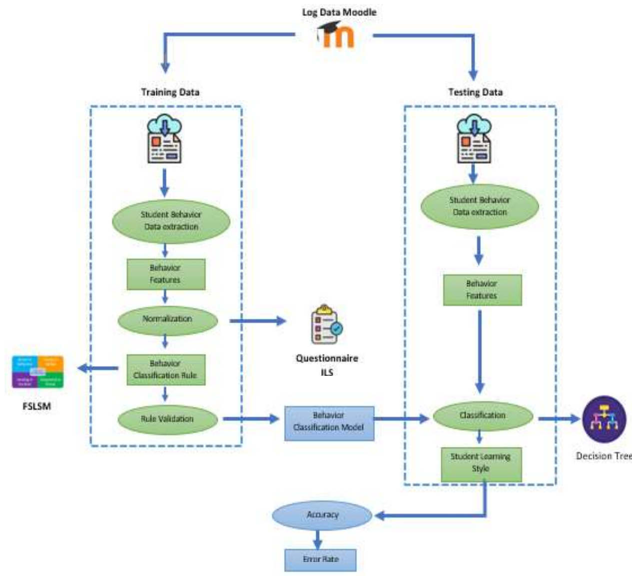

Fig. 2. System Model Architecture

Fig. 2 shows architecture of the model system. The system consists of several sections, including the data collection process (gathering data), pre-processing data, and classification of data.

A. Collect Data

This research is based on data from the Wireless Network Workshop course, which is taught for one semester at the Lamongan State Community Academy. The data collection process in this study is based on the Moodle E-learning log data file at the Lamongan State Community Academy, which is located at the site address http://elearning.aknela.ac.id/moodle/. There were 65 students majoring in Informatics Engineering who were involved in this research. The log data is downloaded from the E-learning system in CSV format. Besides, the data from the questionnaire were also collected from Moodle E-Learning. Fig. 3 shows the course view of the e-learning web.

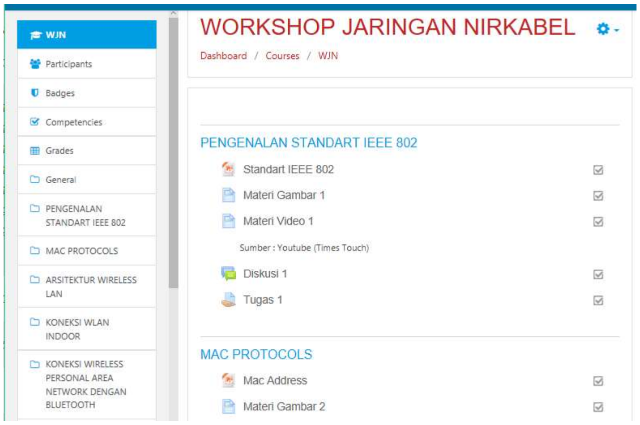

Fig. 3. Display of Moodle E-Learning

B. Preprocessing Data

Process/steps taken to make raw data into quality data. There are several processes in the preprocessing of data, that is:

1) Student Behavior Data Extraction

Data extraction is the process of obtaining information from the Moodle E-Learning log file. Each event log in the logs has nine attributes: time, full user name, affected user, event context, component, event name, description, origin, IP address. In this research, we only focus on the attributes of the full user name, event context, event name and description. The event context and event name attributes represent actions taken by students on various items that can be accessed from Moodle such as chat, assignments, course content, forum discussions.

Fig. 4 shows the output of the log data in Moodle.

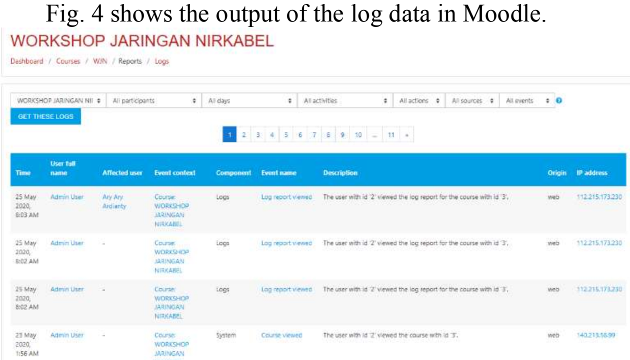

Fig. 4. Display Moodle Data Log

2) Behavior Feature

The process of selecting features from data stored in the Logs file that correspond to the FSLSM learning style dimensions. After students finish learning on E-Learning, all student activities on the courses in E-Learning will automatically be stored in the Moodle log data. For research purposes, we only extract the log data results based on the required attributes by selecting only a few features, namely the full user name, event context, event name, as shown in Fig. 5.

User full name	Event context	Event name
Mukhammad Malik Fajar	URL: Run Linux from Web Browser	Course activity completion updated
Mukhammad Malik Fajar	Course: WORKSHOP JARINGAN NIRKABEL	Course viewed
Mukhammad Malik Fajar	URL: Run Linux from Web Browser	Course module viewed
Mukhammad Malik Fajar	Course: WORKSHOP JARINGAN NIRKABEL	Course viewed

Fig. 5. Moodle Log Data Extraction Results

3) Normalization

The process of normalization is the organization of student behavior data that is collected and compared with the results of the ILS questionnaire, then finding relationships to represent them in a format suitable for creating Behavior Classification Rules. The ILS questionnaire was filled out based on the Felder and Silverman learning style questionnaire on the website https://www.webtools.ncsu.edu/learningstyles/. The questionnaire consists of 44 questions [14].

4) Behavior Classification Rule

Behavior Classification Rule is a rule for classifying learning styles based on student behavior. The features were mapped based on their dimensions according to the Felder-Silverman Learning Style Model[15][20]. Based on the classification rules of the learning style table, we can get a pattern to shape a dataset based on learning objects which are used as FSLSM-related attributes/features. Where the learning style based on FSLSM is divided into 4 dimensions, namely processing (active & reflective), input (verbal & visual), perception (sensor & intuitive) and understanding (sequential & global). Thus it can be said that the learning object forum, demo and chat are part of the processing dimension. In contrast, the learning object text, picture, and video include the input dimension as well as the perception dimension based on the example and assignment learning object, the understanding dimension based on learning object navigation and course overview. The value of each attribute is the sum of the visits for each element. Next, determine a label for each student based on their behavior towards e-learning. Labels are determined based on the classification rules for learning styles in TABLE I with weighting.

TABLE I. Rules of Classification Learning Style Base on Student Behavior

Learning Object	Relevant Behavior	Learning Style	Dimension
Forum	post/reply	active	Processing
	view/read	reflective	
Demo	run	active	
	view	reflective	
Chat	post	active	
	view	reflective	
Text	visit	verbal	Input
	no visit	visual	
Power Point	visit	visual	
	no visit	verbal	
Video	visit	visual	
	no visit	verbal	
Picture	visit	visual	
	no visit	verbal	

Example	visit	sensor	Perception
	no visit	intuitive	
Assignment	submit	sensor	
	view	intuitive	
Navigation	navigating linearly	sequential	Understanding
	navigating globally	global	
Course Overview	view	global	
	no view	sequential	

C. Data Classification

Data Classification is the process of classifying/grouping data where the data used has a class or target label. For this analysis, where the classification procedure uses a supervised learning algorithm, the data classification uses the tree method, namely the Decision Tree and the Ensemble Process [21], respectively, the Gradient Boosted Tree. Decision Tree is defined as a classifier in the form of a data structure for analyzing, designing, and deciding specific patterns. The decision tree starts the test from the root of the tree. Then, the test moves through the tree to a leaf node. Using the pruning process to untie the tree and break leaf nodes with a small number of error points or a certain proportion of the complete training set [17].

The steps in the decision tree begin with moving the training data to a tree, the tree consists of nodes, to determine the location of each node using the Entropy E value specified in formula (1).

$$\mathrm{Entropy}(S) = \sum_{j=1}^{k} -p_j \log_2 p_j \qquad (1)$$

- S is the set (dataset) of cases
- k is the number of S partitions
- Pj is the probability that Sum (Yes) divided by Total Cases

In the decision tree, It can be determined using the benefit value of each attribute when selecting the appropriate segmentation element from the data attribute. It depends on the different meanings this variable takes. This measure is based on research [18].

$$G(S,Q) = E(S) - \sum_{i=1}^{k} p_i \, E(S,Q_i) \qquad (2)$$

- S is entropy
- G is the gain
- Q is an attribute
- pi is the set of values

III. Experiment And Result

The experiment in this analysis is the stage of implementation of a system based system built in the section on system design and result.

A. Dataset Implementation

To get training data that is from the log data in Moodle, which is selected according to the FSLSM model. Therefore the results from the data from the moodle log will be selected and used as the dataset attribute. Furthermore, the labeling process is based on the FSLSM classification rules [15] [20] in TABLE I. Each attribute is the number of total student visits in each learning object. Students are said to be in class processing if they access forum, chat, and demo learning objects. Input class when accessing learning objects text, video, picture, and ppt. Perception class if it obtains the example and assignment learning object. Class understanding when accessing the course overview and navigation learning object. In Fig. 5 below is a mapping per student of the dimensions of learning styles.

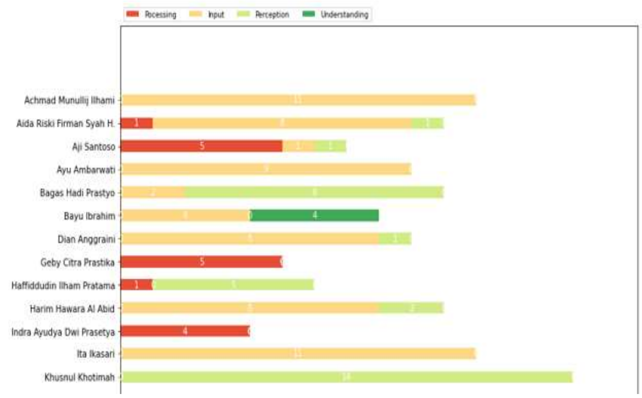

Fig. 5. Mapping students to the dimensions Of learning styles

Fig. 5 shows the results of the student mapping of the learning style dimensions. It can be seen in Figure 5 that the processing dimension is red, the input dimension is orange, the perception dimension is light green, and the understanding dimension is dark green. So that based on the mapping plot, it will be seen that the students are inclined towards the learning style dimension, which is based on the most significant number of dimensions per student. The results of the mapping can be used as class labels in the dataset.

Besides, students also filled out an ILS questionnaire. Each object received a label that matched the student's learning style - identified through the ILS questionnaire [19].

The label of the behavior results and the results of the ILS questionnaire will be compared. The labeled dataset will be used as training data.

Student	Forum	Demo	Chat	Text	PowerPoi	Video	Picture	Example	Assignment	Navigation	Course Overview	Class_dataset
Achmad Munullij Ilhar	0	0	0	1	5	2	3	0	0	0	0	input
Aida Riski Firman Syah	1	0	0	1	1	4	2	0	1	0	0	input
Aji Santoso	3	0	2	0	1	0	0	1	0	0	0	processing
Ayu Ambarwati	0	0	0	0	4	1	4	0	0	0	0	input
Bagas Hadi Prastyo	0	0	0	0	1	0	1	3	5	0	0	perception
Bayu Ibrahim	0	0	0	1	1	1	1	0	0	2	2	input
Dian Anggraini Indah	0	0	0	0	3	4	1	0	1	0	0	input
Geby Citra Prastika	3	1	1	0	0	0	0	0	0	0	0	processing
Haffiddudin Ilham Pra	0	1	0	0	0	0	0	5	0	0	0	perception
Harim Hawara Al Abid	0	0	0	0	4	1	3	0	2	0	0	input
Indra Ayudya Dwi Pras	2	1	1	0	0	0	0	0	0	0	0	processing
Ita Ikasari	0	0	0	1	1	6	3	0	0	0	0	input

Fig 6. Overview of the dataset

Fig. 6 shows a dataset consisting of several attributes: forum, demo, chat, text, powerpoint, video, picture, example, assignment, navigation, and course overview. The value of each attribute is the total student visits to each learning object in e-learning. As for the class/label, the weighted value of the cumulative visits for each category is calculated based on the FSLSM definition following the classification rules for the learning style in TABLE I.

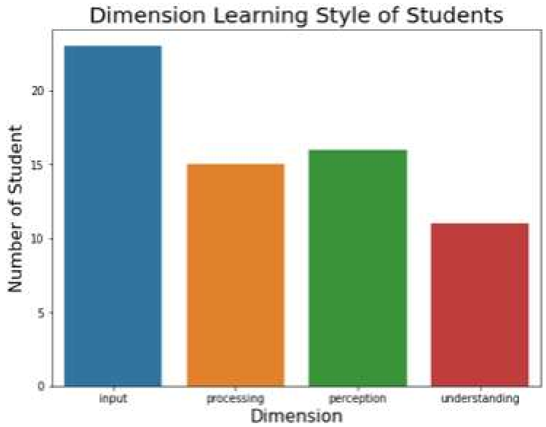

Fig. 7. Dimensions of Student Learning Style Results

Fig. 7 is a plot of the dimensions of student learning styles based on the dataset. It can be seen in the image based on the experimental results that the highest dimensions of students are input (visual & verbal) by 23 students, perception (sensor & intuitive) dimensions by 16 students, processing (active & reflective) by 15 students, and understanding dimensions (sequential & global).) of 11 students.

B. Classification Experiments

In this study, the classification process uses the Decision Tree algorithm and uses the ensemble method, namely the Gradient Boosted Tree. The classification process uses the Rapid Miner tool with 10 Fold Cross-Validation. Cross-Validation is a statistical method that can be used to evaluate the performance of a model or algorithm where data is separated into two subsets, namely learning process data and validation/evaluation data.

The classification model using a decision tree algorithm from the experimental results is illustrated in Fig. 8:

```
Assignment > 2.500: perception {input=0, processing=0, perception=12, understanding=0}
Assignment ≤ 2.500
|   Example > 2: perception {input=0, processing=0, perception=3, understanding=0}
|   Example ≤ 2
|   |   Chat > 0.500: processing {input=0, processing=12, perception=0, understanding=0}
|   |   Chat ≤ 0.500
|   |   |   Forum  > 2: processing {input=0, processing=2, perception=0, understanding=0}
|   |   |   Forum  ≤ 2
|   |   |   |   Demo > 0.500: input {input=1, processing=1, perception=1, understanding=0}
|   |   |   |   Demo ≤ 0.500
|   |   |   |   |   PowerPoint > 0.500: input {input=17, processing=0, perception=0, understanding=0}
|   |   |   |   |   PowerPoint ≤ 0.500
|   |   |   |   |   |   Text  > 2.500: input {input=5, processing=0, perception=0, understanding=0}
|   |   |   |   |   |   Text  ≤ 2.500: understanding {input=0, processing=0, perception=0, understanding=11}
```

Fig. 8. Description of The Model Tree

C. Measurement of Experiment

To assess the accuracy of our method, we use formula (3) and (4) proposed by Ling Xiao et al. [11] and Quang Dung et al. [16].

$$\text{Precision} = \frac{\sum_{1}^{n} \text{Sim}\ (\text{LS}_{\text{determined,}}\ \text{LS}_{\text{ILS}})}{\text{n}} \tag{3}$$

Where Sim is equal to 1 if the value obtained by our method with ILS is equal, 0 if opposite, 0.5 if one is neutral and the other is an extreme value. At the same time, n is the number of students.

$$\text{Accuracy} = \frac{\text{number of predictions correct}}{\text{total number of predictions}} \tag{4}$$

Comparison of the precision and accuracy of the Decision Tree and Gradient Boosted Tree algorithms:

TABLE II. RESULT OF COMPARISON

Algorithm	**Precision**				**Accuracy**
	Processings (active/ reflective)	***Input (visual/ verbal)***	***Perception (sensor/ intuitive)***	***Understanding (Sequential/ Global)***	
Decision Tree	92.86%	86.36%	100.00%	64.29%	85.71%
Gradient Boosted Tree	88.24%	87.50%	80.00%	88.89%	85.95%

Based on the measurements of TABLE II using Rapid Miner, we can conclude that the classification method using the Gradient Boosted Tree ensemble method obtains higher accuracy of 85.95% compared to the Decision Tree. Comparing the classification of the two algorithms, the Gradient Boosted Tree achieves the most elevated of 85.95% and the Decision Tree yields 85.71%.

IV. CONCLUSION

In this paper, we present a new, literature-based method for estimating automatic and dynamic learning styles of learners. Automatically based on the ILS questionnaire and dynamically based on the analysis of student behavior about LMS. Moodle was the LMS used for research because Moodle already had log data that could be used to analyze student behavior towards LMS. Learning styles are generated from total visits to learning objects in e-learning. The resulting learning style results are used as labels on the training data. The proposed method refers to the four dimensions of FSLSM, namely Process (Active / Reflective), Input (Visual / Verbal), Perception (Sensor / Intuitive) and Understanding (Sequential / Global). The method used for classification is by using a decision tree, and ensemble Gradient Boosted Tree methods. We have evaluated this model using the Tree method in the Rapid Miner tool, where there are two types of trials, namely Decision Tree and Gradient Boosted Tree. The final results prove that the tree classifier using the ensemble method can better detect and predict learning styles. This is because the ensemble method combines several single classifications to obtain a more accurate classification model. The Ensemble method improves the accuracy of a single classifier by training several different classifiers. The ensemble method produces 85.95% accuracy, while the processing dimension is 88.24% accurate, the perception dimension 80.00%, the input 87.50%, and the understanding 88.89%. The results are fairly satisfactory.

Future work is to design effective adaptive systems by integrating classifiers into LMS for automatic detection of student learning styles, and we will compare the performance of the decision tree with other classified algorithms, such as Naïve Bayes, Neural Network.

ACKNOWLEDGMENT

The author would like to thank the Lamongan State Community Academy for supporting and collaborating in this research.

REFERENCES

[1] U. K. Mothukuri, U. K. Mothukuri , B. V. Reddy , P. N. Reddy , S. Gutti and et al, "Improvisation of Learning Experience Using Learning Analytics in Elearning," in E-Learning & E-Learning Technologies (ELELTECH), IEEE, 2017.

[2] R. R. Estacio and R. C. Raga, "Analyzing students online learning behavior in blended courses using Moodle," Asian Association of Open Universities Journal, Vol. 12 No. 1, pp. 52-68, April 2017.

[3] D. Susanto, S. Irdoni, M. U. H. A. Rasyid, "Attendance Report Plugin for E-Learning Applications in PENS," in International Electronics Symposium on Knowledge Creation and Intelligent Computing (IES-KCIC), IEEE, 2017.

[4] M. S. Hasibuan, L. E. Nugroho and P. I. Santoso , "Detecting Learning Style Based on Level of Knowledge," in Third International Conference on Informatics and Computing (ICIC), IEEE, 2018.

[5] H. T. Binh and B. T. Duy, " Predicting Students' performance based on Learning Style by using Artificial Neural Networks," in International Conference on Knowledge and Systems Engineering, 2017.

[6] M. S. Hasibuan and L. E. Nugroho, "Detecting Learning Style Using Hybrid Model," in e-Learning, e-Management and e-Services (IC3e), IEEE, 2016.

[7] J. Wang, "Research on Online Learning Behavior Analysis Model in Big Data Environment," EURASIA Journal of Mathematics Science and Technology Education, vol. 3, pp. 5676-5684, 2017.

[8] A. Quafae, A. Yasser, O. Lahcen and A. Youssouf, "Integrating Web Usage Mining for an Automatic Learner Profile Detection," in Intelligent Systems and Computer Vision (ISCV), IEEE, 2018.

[9] A. Quafae, A. Yasser, O. Lahcen and A. Youssouf, "A Hybrid Machine Learning Approach to Predict Learning Styles in Adaptive E-Learning System," in International Conference on Advanced Intelligent Systems for Sustainable Development, January 2019.

[10] M. Abdullah, A. Alqahtani, J. Aljabri, R. Altowirg, and R. Fallatah, "Learning Style Classification Based on Student's Behavior in Moodle Learning Management System," in Transaction on Machine Learning and Artificial Intelligence, vol. 3, 2015

[11] L.X. Li and S. S. A. Rahman, "Students' learning style detection using tree augmented naive Bayes," in Royal Society Open Science, 2018.

[12] U. F. Alias, N. B. Ahmad, and S. Hasan, "Student Behavior Analysis Using Self-Organizing Map Clustering Technique," in ARPN Journal of Engineering and Applied Sciences, vol.10, no.23, December 2015.

[13] Logs in Moodle. [Online]. Available: https://docs.moodle.org/38/en/Logs. accessed May 22, 2020.

[14] R. M. Felder, and B.A. Soloman, Index of Learning Styles Quistionnaire,https://www.webtools.ncsu.edu/learningstyles/,2004, accessed May 22, 2020.

[15] R. M. Felder, "Learning and Teaching Styles in Engineering Education," in Journal of Engineering Education, 78(7):674-681, January 1988.

[16] Q. D. Pham and A. M. Florea, "A Method For Detection Of Learning Styles In Learning Management Systems," in UPB Scientific Bulletin, Series C: Electrical Engineering, 75(4):3-12, January 2013.

[17] A. Xavier, A. Jaimes, N. Oliver and J. M. Pujol, "Data Mining Methods for Recommender Systems, " Recommender Systems Handbook, Springer US, 2011. 39-71.

[18] B. Hssina, A. Merbouha, and B. Bouikhalene, "Predicting Learners' Performance in an E-Learning Platform Based on Decision Tree Analysis, " in International Arab Conference on Information Technology (ACIT), 2016.

[19] L. D Ferreira, G.l Spadon, A. C. Carvalho and J. F. Rodrigues,"A comparative analysis of the automatic modeling of Learning Styles through Machine Learning techniques," in IEEE Frontiers in Education Conference (FIE), IEEE,2018.

[20] R. R. Maaliw, "Classification of Learning Styles in Virtual Learning Environment using Data Mining: A Basis for Adaptive Course Design"in International Research Journal of Engineering and Technology (IRJET), vol.3,no.7, July 2016.

[21] G. Seni and J. Elder, "Ensemble Methods in Data Mining: Improving Accuracy Through Combining Predictions," in IEEE , Morgan & Claypool, 2010.