Naïve Bayes for Detecting Student’s Learning Style Using Felder-Silverman Index

Nurul Hidayat¹, Lasmedi Afuan²

¹²Department of Informatics, Faculty of Engineering, Universitas Jenderal Soedirman, Indonesia

¹nurul@unsoed.ac.id, ²lasmedi.afuan@unsoed.ac.id

Abstract - This paper focuses on detecting student learning styles using the Felder-Silverman Index Learning Style (FSLM). Providing Adaptivity based on learning styles can support students and make the learning process easier for them. However, the student learning styles need to be identified and understood to provide the appropriate adaptability. In this case, we use a questionnaire instrument to detect student’s learning styles. This paper analyses of students from Professional Education Teacher (PPG) at the Ministry of Research, Technology, and Higher Education (Kemenristek DIKTI). The results show that 1998 students who filled out the questionnaire obtained the following conclusions for each zone with a balanced learning style about 29.9% for dimension processing, 34.78% for input dimension, and 36.98% for understanding dimension. However, most students have a moderate sensing learning style with 31.13% for each zone for the dimension of perception. This research contributes to some areas, such as providing FSLSM learning style with a large dataset and capturing students' learning styles based on four dimensions.

Keywords: Index Learning Style, E-learning, learning style, Felder-Silverman model, questionnaire

I. INTRODUCTION

The approach to detect the learning styles can be divided into two methods, namely (1) static detection based on the questionnaire and (2) dynamic detection through learning behaviour [1]. Learning styles are a set of characteristic, emotional, cognitive, and physiological factors that has function as relatively stable indicators of how students understand, interact with and respond to the learning environment [2]. Studies on the influence of learning styles are influenced mainly by students' learning attitudes, levels of satisfaction, and academic achievement in an online learning environment [2]. Learning styles can significantly affect learning attitudes in the educational environment. On the contrary, when students' learning styles do not match, learning effectiveness is reduced [1]. Some of the researchers consider learning styles in the development of e-learning systems. This learning style aims to maintain student motivation to take part in the learning process more effectively [3-4].

According to [5], the advantages in development adaptive e-learning are possible to (1) overcome the limited number of teaching hours, (2) students can learn by using an individual approach, (3) increasing mastery of students' understanding of learning material, and (4) therefore the objectives learning can be achieved as stipulated conditions. The adaptive referred to in this e-learning development research is an adaptation to visual, audio, and kinesthetic learning styles. Teaching materials or known as material contents, are developed by considering students' learning styles. Therefore, the material and information conveyed can be received definitely.

Numerous research on e-Learning personalization uses Felder-Silverman Learning Model (FSLM) as an indicator of learning styles. For example, this research is conducted to provide different learning needs with varying styles of learning using adaptive hypermedia and a recommendation system. Usually, the research methodology used in previous research decomposes into two main phases. First, the system will create a learning profile by assessing student learning styles using a learning style index questionnaire. Second, after combining every aspect, students are given a suitable learning environment in e-Learning [6-7]. Another example is FSLM, which is used to create teaching strategies combined with suitable electronic media such as wikis, videos, emails, and others.

Over the last few years, personalization has already addressed student adaptation during e-learning process-based [8-9]. Students, furthermore, often get difficulty with a large amount of information that might be related to their interests. The way in presenting the learning material (e.g., learning objects only) regarding learning styles is one of the essential problems for the recommended learning system [10]. The single technique to solve this problem is creating a learner model considered a core component in an intelligent or adaptive learner recommendation system. The learner model represents many learner features, such as
knowledge and learning styles, so that it can be accessed to offer adaptation [11].

Students must be considered as the core of the teaching situation to encourage integration into the learning process. Adaptive learning uses techniques to interpret student activities based on domain-specific models, infer students’ needs from interpreted activities, accurately represent the needs of related models, and act on adaptive student profiles to facilitate the learning process dynamically. Consequently, most of the learning programs try to overcome all suitable approaches. Adaptive learning systems can enhance the individualization of student learning [12] by changing content and address for students based on their learning profiles, as shown in Fig. 1. Student profiles include some data such as personal information, knowledge, and learning styles. Two main approaches for detecting learning styles are explicit modelling (questionnaire-based) and implicit modelling (literature-based).

The explicit modelling technique represents each student's learning characteristics and needs based on the data obtained by requiring each student to fill out a learning style questionnaire. Examples of systems that use this explicit modelling method are CS383 [13] INSPIRE [11] and iWeaver [14]. The implicit modelling approach means that the adaptive system continually updates student models by monitoring interactions with systems; examples include the Arthur system [15] dan Protus 2.0 [6].

The Index of Learning Style (ILS) is an instrument used to identify learning styles based on the FSLM. Moreover, ILS consists of 44 statements representing each dimension of FSLM, which means 11 statements for each dimension. After assessing the questionnaires, students will be shown the tendencies of their learning style. There will be sixteen learning styles that allow combinations [16]. Several studies have been conducted to analyze and measure ILS in the English version. For example, Table I shows previous research from the Questionnaire about Index of Learning Style [16].

Generally, every learning style model has its own assessment tool in the form of a questionnaire. This learning style provides a variety of questions about personality, attitudes and learning behaviour. Learning style inventory helps people be more aware of their learning style and realize that they also have limitations. Hence, the students should not label his style too narrow [17]. There are several theories related to learning styles, such as Felder-Silverman's learning style model (Felder 88), Honey and Mumford, Kolb learning style model.

Each of those learning styles proposes a description and classification of different learning styles. In this study, we only focus on the Felder Silverman learning style model (FSLM). Most other learning style models classify learners, while FSLM describes more the learning style of a student in Australia in detail, distinguishing between preferences on four dimensions [18] (active/reflective, sensing/intuitive, visual/verbal and sequential/global). Therefore, each learner has a preference for each of these four dimensions [19].

In addition, the proposed Online Network Learning System by the Ministry of Research, Technology and Higher Education (KEMENRISTEK DIKTI) combines two approaches. The questionnaire of FSLM [20] is used to determine the learner's initial learning style and user preferences to initialize the learner's profile. During system use, student profiles are dynamically adapted based on user behaviour (i.e., interactions with the system), knowledge and performance in learning. This paper focuses on the initialization of adaptive learner profiles by using dynamic variants of the FSLM questionnaire.

The use of the Naïve Bayes method in this study due to has some advantages. First, Naïve Bayes can process a real-time dataset quickly and save processing time effectively. Second, Naïve Bayes is a feasible method for solving multi-class prediction time. Since we use four different dimensions related to learning preference, this method is able to cover the multi-pairwise dimension comparison. Third, the Naïve Bayes method can perform better than other methods like Decision Tree and Random Forrest because it requires less training dataset. Four, the Naïve Bayes is acceptable and suitable for categorizing input variables rather than numerical variables.

Fig. 1 The classic model of student profile adaptation
The main contributions of this paper are in the form of three things, they are such as a) Analytic to build an adaptive learner profile during registration based on Felder-Silverman learning style model; b) An empirical study to determine the order of questions for each of four dimensions from Felder Silverman's learning style questionnaire; c) Algorithms was built based on the ranking of the questions, and it is used to dynamically calculate the user's initial learning style through the questionnaire; d) Capture student learning styles based on Indonesian time zones. An innovative feature of this algorithm is its ability to determine learners' learning styles in each dimension from user responses to a few questions only from the questionnaire. Therefore, it can set aside the users time and effort from answering all 44 questions from the Felder-Silverman learning style questionnaire.

II. METHOD

The block diagram in Fig. 2 illustrates the proposed method for initializing students' adaptive profiles based on a dynamic learning style questionnaire. Student profiles are included in the personal details of students, which are collected from students during the registration process. After the registration process, students are asked to fill in the ILS questionnaire described in the previous section. The algorithm for creating adaptive learner profiles is described in some steps below:

1. Step 1: (Registration Initialization). students should register through the portal of e-learning provided by PPG (Professional Education of Teacher) of SPADA (Online Network Learning System) from the Ministry of Research Technology and Higher Education (Kemenristekdikti) before joining the system. During the registration process, personal data such as personal name, e-mail address and password are gathered.

2. Step 2: (Fill the ILS questionnaire). After the registration step, the students require to take the ILS questionnaire. When students answer the questionnaire, the system dynamically calculates student learning styles before the dimension by counting the number of answers "a" and the number of answers "b". When the number "a" (or "b") reaches the score of 7 (i.e., 60% of 11 questions) in one dimension, the system skips the remaining questions for that dimension and moves to the first question from the next dimension.

3. Step 3: (Calculate the value of learning styles for each dimension). Calculate the learning styles for each dimension as percentages "a" and "b". For example, a person might have 60% visual and 40% verbal in recording dimension information.

4. Step 4: (Saving values the learning style in student profiles). The initial learning styles, which are calculated through the ILS questionnaire, are distributed in the student profile database.

A. Model of Felder-Silverman Index Learning Style (FSLSM)

Learner learning styles have been identified as essential factors that influence the learning process. Learning style is the most significant parameter for personalization. Students have differences in the way of understanding, processing and receiving information. Students are considered to have their learning styles based on how they process and organize information.

Fig. 2 shows four dimensions of the FSLSM [20] associated with information processing, understanding, input, and perception. Each of these dimensions is marked by a pair of X/Y (i.e., active/reflective, sequential/global, visual/verbal, and sensing / intuitive), which means that learners' learning style is X or Y extent. One example is in the dimension of information processing active or reflective user learning styles to some extent. In the dimension of information input, the users can be visual or verbal to some extent. FSLSM is considered the most stable and appropriate learning style model for adaptive hypermedia learning systems [21].

B. Experimental Setting

1) Dataset: This study uses ILS data questionnaires on English subjects. The sample used for testing is Slovin equation (1), such as:

\[ n = \frac{N}{1+Ne^2} \]
Based on (1), the ILS questionnaire's data population in the amount of 2104 at an error rate of 0.5% obtained a sample of 1998 students. The students who become the research sample then fill out the questionnaire, which was conducted online.

2) Collecting Data of ILS Questionnaire: ILS data collection is carried out by asking students to complete a questionnaire. The questionnaire consisted of four dimensions, each dimension composed of 11 questions which became the FSLSM learning style model. The display of questionnaire entries is presented in Fig. 3.

Students are asked to choose one of two answer choices for each question related to learning style. There are 44 questions raised and divided into four dimensions, and each consists of 11 questions. If students have answered all questions, the ILS application will display predicted learning styles from students, as shown in Fig. 4.

3) Classification of Learning Style Models with Naïve Bayes: Classification is a process of grouping data based on specific characteristics into classes that have been determined. Some classification methods often used include Naïve Bayes, Decision Tree J48, Bayes Net and Random Forest. This study is using Naïve Bayes because it has a higher level of accuracy compared to other models. Bayes theorem is shown in (2).

\[
P(C|X) = \frac{P(X|C) \cdot P(C)}{P(X)}
\]

where:
- \( P(X) \) is the prior probability of predictor
- \( P(C) \) is the prior probability of class
- \( P(C|X) \) is the posterior probability of class (C, target) given predictor (X, attributes)
- \( P(X|C) \) is the likelihood which is the probability of predictor given class.

![Fig. 3 The display of ILS questionnaire](image-url)
Looking for $P(C|X)$ with a maximum value, as well as $P(X|C) \cdot P(C)$ is also in a maximum value.

If it is given to the $k$ attributes that are free (independence), the probability value can be given as in (3).

$$P(x_1, \ldots, x_k|C) = P(x_1|C) \cdot \ldots \cdot P(x_k|C)$$  \hspace{1cm} (3)

If the attribute of $i$ is discrete, then $P(x_i|C)$ is estimated as the relative frequency of the sample, which has a value of $x_i$ as $i$ attribute in class $C$, but if attribute of $i$ is continuous, then $P(x_i|C)$ is estimated with the Gauss density function, described in (4).

$$f(x) = \frac{1}{\sqrt{2\pi \sigma}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$  \hspace{1cm} (4)

$\mu = \text{mean}$, $e = \text{rate of error sample}$ and $\sigma = \text{standard deviation}$.

Measurement of assessment and testing in classification models can be determined by using several techniques, including Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), Relative Absolute Error (RAE), Root Relative Squared Error (RRSE).

III. RESULTS AND DISCUSSION

From the data collection stage, a recapitulation procedure was carried out for each student. The answers in each group classified according to the dimensions of each question. The algorithm is used to determine the sum of the values for each dimension and answer falls. The value of the dimension category, as shown in Fig. 5.

Description of result each stage:
- If the value score of the scale is 1-3, the result is good on the two-dimensional level
- If the value score of the scale is 5-7. In that case the result is moderate preferences for the one dimension and will learn more efficiently in a learning environment that supports the dimension
- If the value score of scale is 9-11, the result is very strong for one scale dimension. The experience of learning difficulties in an environment that does not support that preference.

The result of grouping students’ answers on each dimension is presented in Tables I - IV. Algorithm for determining dimensions shown in Fig. 5.
Table I presented data that there are 833 students entered into the moderate active dimension attribute, 34 students entered into the moderate reflective dimension attribute, 253 students entered into the strong, active dimension attribute, three students entered into the strong reflective dimension attribute, and 872 students entered into the balanced attribute. Besides, Table II presented data 891 students entered into the moderate sensing dimension attribute, 15 students entered into the moderate intuitive dimension attribute, 523 students entered into the strong sensing dimension attribute, 0 students entered into the strong intuitive dimension attribute, and there are 566 students entered into the balanced attribute.
Table III presented data about two students who entered into the strong global attribute, 59 students entered into the moderate global dimension attribute, and 1087 students entered into the balanced attribute. Likewise Table IV is included in the dimension table of V/V and presents data that there are 589 students entered into the moderate visual dimension attribute, and 1004 students entered into the balanced attribute. From four tables above, a recapitulation is performed as shown in Table V – IX.

Based on Table V, it shows the dimensions of A / R for all regions (WIB, WITA, and WIT) that tend towards a balanced learning style with a value of 598, 257, and 17. Then followed by moderate active and strong, active attributes. It can be seen that S/I dimensions for all regions (WIB, WITA, and WIT) tend to moderate sensing learning styles with values of 622, 252, and 17, like in Table VI. Furthermore, it is followed by the balance attribute and the strong sensing attribute.

Table VII shows that the V/V dimensions for all regions (WIB, WITA, and WIT) have a tendency for balanced learning styles with values of 695, 297, and 12. Furthermore, it is followed by moderate active and strong active attributes. In Table VIII, the WIB region tends to learn styles that are more balanced than the other areas. Besides, Table VIII states that the dimensions of S/G for all regions (WIB, WITA, and WIT) have a tendency of a balanced learning style with values of 739, 329, and 19. Furthermore, it is followed by moderate active and strong active attributes. In general, the WIB region tends to balance.

Fig. 6 shows that overall the Processing dimension has a stronger balance attribute than other attributes. Especially for WIB regions, it has a higher balance value compared to other regions. Overall of the perception dimension showed in Fig. 9.

Fig. 7 shows that overall the perception dimension has a moderate sensing attribute that is stronger than the other attributes. And especially for WIB regions, it has a higher Moderate sensing value compared to other regions. While Fig. 8 shows that overall the input dimension has a stronger balance attribute compared to other attributes. Especially for WIB regions, it has a higher balance value compared to other regions. Overall the understanding dimension has a stronger balance attribute compared to other attributes as showed in Fig. 9. Especially for WIB regions, it has a higher balance value compared to other regions.

Comparisons with several models such as Bayes Net, Random Forest and Trees J48 are presented in Table VII with the classification calculation process using the help of weka software tools. The result of the classification using the Naïve Bayesian model is described in Table VIII, where the four dimensions are intuitive sensing, visual verbal, global sequential, and active reflective. The results obtained from weka software using the naïve Bayes model can be described in Table IX.

| TABLE V |
| ILS QUESTIONNAIRE SUMMARY FOR A/R DIMENSION |
| Zone | Balance | Processing |
| | | Active | Reflective | Strong |
| | | | Active | Reflective |
| WIB | 598 | 584 | 23 | 170 | 1 |
| WITA | 257 | 238 | 11 | 82 | 2 |
| WIT | 17 | 11 | 0 | 1 | 0 |

| TABLE VI |
| ILS QUESTIONNAIRE SUMMARY FOR S/I DIMENSION |
| Zone | Balance | Perception |
| | | Sensing | Intuitive | Sensing | Intuitive |
| WIB | 380 | 622 | 10 | 364 | 0 |
| WITA | 176 | 252 | 5 | 157 | 0 |
| WIT | 10 | 17 | 0 | 2 | 0 |

| TABLE VII |
| ILS QUESTIONNAIRE SUMMARY FOR V/V DIMENSION |
| Zone | Balance | Input |
| | | Visual | Verbal |
| | | Moderate | Strong |
| WIB | 695 | 400 | 63 | 214 | 4 |
| WITA | 297 | 179 | 23 | 89 | 2 |
| WIT | 12 | 10 | 4 | 3 | 0 |

| TABLE VIII |
| ILS QUESTIONNAIRE SUMMARY FOR S/G DIMENSION |
| Zone | Balance | Understanding |
| | | Moderate | Global | Strong |
| | | Sequential | Global | Sequential | Global |
| WIB | 739 | 489 | 42 | 105 | 1 |
| WITA | 329 | 201 | 17 | 42 | 1 |
| WIT | 19 | 9 | 0 | 1 | 0 |
Table IX shows the total data used the total of 1998 data. The correctly Classified Instances value is 54.5%, which means that the true value during the classification process is 54.54%. As for Incorrectly Classified Instances, the value is 45.45%, which means the misclassification rate is 45.45%. In the naïve Bayes model, most of the data can be recognized by 908 data.

The measurement of classification values using the MAE approach is in the number of 0.26, RMSE is in the number of 0.36, RAE and RRSE have the same value as 100%. Table X shows the comparison of each dimension of learning styles and comparison of the classification models used. The classification model used for comparison is Decision Tree J48, Bayes Net and Random Forest. The value used as a comparison is RAE. In the Naïve Bayes model, the RAE value obtained in the four dimensions is 100%, while the J48 decision tree model has an average value of 98%, Bayes Net Model is around 99%, and the random forest is 99%. Based on this result, we postulate that the naïve Bayes model can classify learning styles for each dimension.
IV. CONCLUSION

This study has detected the learning styles of students by using questionnaires combined with the FSLSM method. The 1998 students who filled out the questionnaire obtained the following conclusions overall for each zone with a balanced learning style with 29.9% for dimension processing, 34.78% for input dimension, and 36.98% for understanding dimension. However, most students have a moderate sensing learning style with 31.13% for each zone for the dimension of perception. For the next research step, a student activity-based detection study was conducted using a Log file in LMS. Capture student learning styles based on Indonesian time zones. This paper provided new findings regarding comparing the four different methods (Naïve Bayes, Decision Tree J48, Bayes Net, and Random Forest). Our experiment shows that the Naïve Bayes method is the most feasible method to classify the learning styles for four different dimensions. Our study contributes to the researchers in the state of the arts of implementing the FSLSM method with real empirical settings. For future work, we recommend using several other methods like linear regression, Dimensionality Reduction Algorithm (RDA), and Gradient Boosting Algorithm (GBA) to obtain an in-depth novelty in comparing student’s learning styles.

ACKNOWLEDGMENT

We would like to thank LPPM UNSOED (1072/23/HK.02/2021) for research funding support and thank colleague supporting this research.

REFERENCES

