JURNAL INFOTEL
Informatics - Telecommunication - Electronics
Website: http://ejournal.st3telkom.ac.id/index.php/infotel
ISSN: 2085-3688; e-ISSN: 2460-0997

Detection of learning styles with prior knowledge data using the SVM, K-NN, and naïve bayes algorithms

Muhammad Said Hasibuan[1,*], RZ Abdul Aziz[2]
[1,2]Institut Informatika dan Bisnis Darmajaya
[1,2]Jl. ZA Pagar Alam No. 93 A, Lampung 35136, Indonesia
[*]Corresponding email: msaid@darmajaya.ac.id

Received 14 June 2022, Revised 18 July 2022, Accepted 3 August 2022

Abstract — Data-driven (DD) and literature-based techniques for autonomous learning style detection are the two categories (LB). Both automatic learning style detection techniques offer advantages over traditional learning style detection methods because they leverage external data sources that are more accurate than surveys in conventional styles of detection, such as forums, quizzes, and views of teaching materials. On the other hand, automatic detection results do not always reflect learning styles. this work provides a learning style recognition algorithm that draws on data from the learner's internal source, namely past knowledge, as the proposed method the solve the issues. Prior knowledge is advocated because it is based on the learner's knowledge or skills, which better reflect the learner's traits rather than the learner's dynamic behavior. This research proposes a method for recognizing autonomous learning patterns that rely on prior information. The learning style detection framework is unusual in its three stages: prior knowledge question formulation, prior knowledge measurements, and learning style detection utilizing SVM, Naïve Bayes, and K-Nearest Neighbour (K-NN) classification algorithms. The results showed that Naïve Bayes has an accuracy value of 91.48%, K-NN of 89.39%, and SVM of 87.31%.

Keywords – detecting, learning style, prior knowledge

I. Introduction

A Learning Management System (LMS) combines academic service features such as discussion forums, evaluations, quizzes, and teaching materials into one system [1]. Moodle, Blackboard, and WebCT are some examples of LMS programs. Unfortunately, Personalization elements in LMS' are lacking, and they cannot fit unique learning needs. This flaw has been addressed by incorporating LMS customization capabilities such as learning style identification [2].

Viola and Graf stated that when the learning style is detected, the learner's concentration increases and maximizes the learning process [3]. There are two learning style detection methods: traditional and automatic [4]. However, the conventional learning style detection approach has weaknesses, not only in the time length required but also in terms of its accuracy. The accuracy problem arises because some students complete the questionnaire without understanding its contents [5].

This condition is because filling out the questionnaire requires deep understanding and motivation. An automatic learning style detection mechanism was created to address this flaw. The automatic learning style detection mechanism retrieves learner interaction data. The data include learner interactions with teaching materials, quizzes, discussion forums, and chat activities. All interaction data are recorded automatically so that no additional time is required to detect learning styles, and, compared to conventional methods, more accurate results are produced [5].

However, the accuracy of automatic learning style detection can still be improved. Opportunities for increasing accuracy arise because of interaction data and conditions and because the available teaching materials do not always allow the actual learning style to be determined. For example, if the teaching materials are limited, students may not have the opportunity to encounter their preferred learning style. Conversely, if a wide diversity of teaching materials are provided, students may spend most of their time on activities that will not help to determine their learning styles [6].

The automatic detection approach falls short in the interaction process because it may not accurately reflect learning preferences. Thus, interaction should not be used to increase accuracy as a basis for determining learning styles. Instead, detection accuracy can be improved when the information used is sourced from within the learners themselves, and one of the candidates for this internal information is prior knowledge. Prior knowledge refers to the learner's prior knowledge and skills [7]. Prior knowledge into two types by Hailikari: declarative knowledge and procedural knowledge [7]. Declarative knowledge is the learner's fundamental understanding, also known as 'knowing that,' whereas procedural knowledge is the application of that knowledge, also known as 'knowing how' [8]. The prior knowledge concept discussed by Hailikari *et al.* relates to Bloom's taxonomy. Bloom's taxonomy describes the six levels of learning targets, while prior knowledge shows a mastery of knowledge resulting from learning targeted by Bloom's taxonomy [9]. Several studies on detecting learning styles have been carried out using conventional and automated approaches.

The research by Pantho and Tiantong focused on 1025 student respondents who filled out the VARK questionnaire [10], [11]. They then compared it using the C45 decision trees classification algorithm. Pantho and Tiantong. built 108 rules divided into 22 visual, 24 aural, 37 reading, and 25 kinaesthetic. They then divided learners into three age groups: youth, middle and senior. Their results indicated an accuracy rate of 83.40%. However, this research has not yet provided teaching materials for students.

The research by Crockett *et al.* [12] studied the detection of FSLSM learning styles using the Fuzzy Sugeno method. The respondents were 41 students on a Structured Query Language course. Each learner completed the FSLSM questionnaire and then confirmed it using Fuzzy Sugeno.

Bernard *et al.* [13] used artificial neural network approaches, genetic algorithms, ant colony systems, and particle swarm optimization to identify FSLSM learning styles. Students comprised approximately 75% of the survey respondents. Three metrics emerged from this study: accuracy (ACC), lowest accuracy (LACC), and percentage-matched. This study had an accuracy rate of 80.7%.

The NBTree classification algorithm was used to detect FSLSM learning styles in the study by Abdullah *et al.* [14]. The researchers employed a questionnaire and interactions with 33 King Abdul Aziz campus students who used blackboards in the LMS to detect learning styles. The accuracy of their results was 69.697%.

Kolekar *et al.* [15] studied the detection of FSLSM learning styles using the artificial neural network method among 108 learners who took HTML subjects. The results showed an accuracy rate of 95.93% with 200 iterations.

In another study, Garcia *et al.* [16] adopted the Bayesian Network (BN) technique to determine FSLSM learning styles. FSLSM has three dimensions, according to this study: understanding, perception, and processing. The learning styles of 27 BN students were found to be 77% perception, 63% understanding, and 58% processing.

Ozpolat *et al.* [17] focused on detecting FSLSM learning styles and used the NBTree classification method to detect the learning styles of 25 students. The results were divided into four dimensions: perception (73.3%), understanding (73.3%), processing (70%) and input (53.3%).

This study aims to develop the detection of learning styles by using a prior knowledge approach compared with data-driven and literature-based approaches. The detection process is carried out in several stages: generating prior knowledge, mapping prior knowledge with VARK learning styles, detecting learning styles, and recommending teaching materials

II. Research Method

The following is a learning style detection methodology that utilizes prior knowledge (see Fig. 1):

A. Stage 1: Generating Prior Knowledge

In this step, students who have registered in the course carried out prior knowledge measurements. The measurement process is carried out by asking the learner questions to determine their level of prior knowledge.

The equations used to measure the results are

$$W - Bleu = Bp \times \exp(\sum_{n-1}^{n} W_n \log(WP_n)) \quad (1)$$

where Bp is the brevity penalty obtained from (2).

$$Bp = (\frac{1, s_{a>r_a}}{\exp(a\frac{ra}{sa})}, sa < ra \quad (2)$$

$$Score = W - Bleu \times mm \quad (3)$$

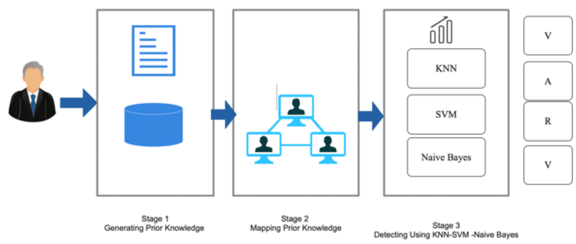

Fig. 1. Systems architecture.

The steps for measuring prior knowledge are as follows:

1) Measure prior knowledge by answering questions given at the beginning of the lesson. The answer of each learner will determine the level of prior knowledge that the learner has.
2) After the prior knowledge level is known, a mapping between prior knowledge and learning styles is carried out. This mapping process produces the learning styles of the learners.

B. Stage 2: Mapping Prior Knowledge

The goal of Mapping Prior Knowledge (MPK) was to highlight the correlation between learning styles and prior knowledge. The association between learning styles and prior knowledge in the mapping is based on past research showing a link between prior knowledge and learning styles [18].

C. Stage 3: Detecting Learning Style

The next stage is to identify learning styles using the Nave Bayes, Support Vector Machine (SVM), and K-Nearest Neighbour classification algorithms (K-NN) [19]–[21].

D. SVM Algorithm

SVM A machine learning approach for making predictions in both classification and regression settings, SVM is a machine learning method. The purpose of SVM is to use the Structural Risk Management (SRM) approach to find the best hyper-lane that divides two classes in the input space [22], [23].

$$(y_1, x_1), \cdot, (y_1, x_1), x \in R^n, y \in [-1, +1] \quad (4)$$

With hyperplane

$$(w, x) + b = 0 \quad (5)$$

$$[(w, q) + b] \geqslant 1 \ for \ y_i = +1 \quad (6)$$

$$[(w, q) + b] \geqslant 1 \ for \ y_i = -1 \quad (7)$$

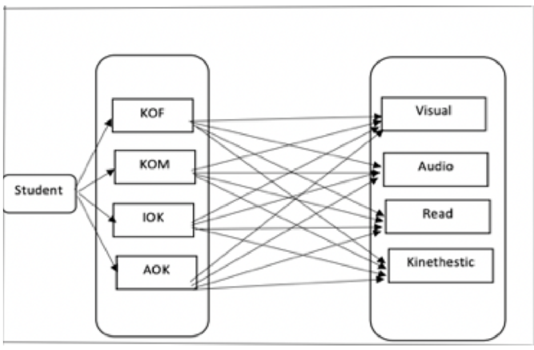

Fig. 2. Prior knowledge mapping.

E. K-NN Algorithm

The K-NN classifies data using the concepts of retrieving by similarity and voting. Voting is then used to determine the final output after the K-NN retrieves the closest K examples to the new one. The value of K significantly influences K-NN accuracy; for example, if we choose a small number for K, other valuable examples can be overlooked, reducing accuracy, whereas a large value for K requires a lot of time and resources. There are various options for selecting the right value of K; the most popular is to compute the square root of the total number of data points.

$$D = \sqrt{\sum_i (x_i - y_i)^2} \quad (8)$$

F. Naïve Bayes Algorithm

The Bayesian classifier is one of the statistical classifiers; this classifier can predict the probability of class membership of a data tuple who will enter a certain class, according to a probability calculation. Bayesian classifiers are based on Bayes' theorem, published by Thomas Bayes in the 18^{th} century. In a comparative study, the classification algorithm is a simple Bayesian or a Naïve Bayes classifier. Naïve Bayes classifiers show high accuracy and speed when applied to large databases.

$$P(A|B) = P(A)P(B)P(B|A) \quad (9)$$

The equation used to measure the accuracy of the detection results is

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad (10)$$

$$Precision = \frac{TP}{(TP + FP)} \quad (11)$$

$$Recall = \frac{TP}{(TP + FN)} \quad (12)$$

III. Results and Discussion

A. Questions Building Stage

The first step is to create a questionnaire to generate prior knowledge. The process of creating a questionnaire containing prior-knowledge-level guidelines, such as Knowledge of Fact (KOF), Knowledge of Meaning (KOM), Integration of Knowledge (IOK), and Application of Knowledge (AOK), involves splitting prior-knowledge level guidelines into four categories [7], [24], [25].

Questions were built using the target knowledge keywords in Table 1. Questions that were built later

Table 1. Prior Knowledge Questionnaire Keywords

Prior Knowledge	Keyword
KOF	Recognising
	Enumerating
	Recalling
	Remembering
KOM	Defining
	Reproducing
	Understand
IOK	Understand Concept
	Classifying
	Comparing
AOK	Problem Solving
	Application
	Producing
	Implementation

needed to be measured for validation and reliability. This research reviewed validation and reliability, as shown in Table 2.

B. Questionnaire Validation Stage

Based on Table 2, the Alpha reliability value is 0.8159, which means that the questions built are reliable because the value is above 0.6. Therefore, the questions can be used for research.

C. Algorithm Implementation Stage

This stage uses the results of the prior knowledge measurement that the learner has with the attached data. Based on the results of prior knowledge generation, a value is generated that is used to determine the level of prior knowledge. Prior knowledge data become inputs for predicting learning styles using the SVM, Naïve Bayes, and K-NN algorithms.

Table 2. Cronbach Alpha

Questionnaire	Alpha	Std Alpha	R (item. total)
P1	0.8166	0.8188	0.3034
P2	0.8086	0.8115	0.3960
P3	0.8120	0.8115	0.3520
P4	0.7953	0.7956	0.5545
P5	0.7953	0.7956	0.5545
P6	0.7953	0.7956	0.5545
P7	0.8050	0.8073	0.4376
P8	0.8062	0.8087	0.4229
P9	0.8098	0.8119	0.3815
P10	0.7925	0.7961	0.5836
P11	0.7910	0.7949	0.5909
P12	0.8000	0.8027	0.4945

Table 3. Dataset of Prior Knowledge

NIS	KOF	KOM	AOK	IOK	Label
6581	**3**	**3**	**2**	**2**	**A**
6606	**4**	**3**	**3**	**0**	**V**
6623	**2**	**3**	**2**	**0**	**A**
6632	**3**	**3**	**2**	**3**	**K**
6553	**4**	**3**	**2**	**3**	**V**
6635	**4**	**3**	**0**	**2**	**V**
6532	**3**	**3**	**2**	**3**	**K**
6570	**3**	**3**	**4**	**3**	**A**
6620	**3**	**2**	**3**	**3**	**V**
6630	**3**	**3**	**3**	**3**	**K**
6538	**3**	**3**	**3**	**2**	**R**
6604	**4**	**3**	**2**	**2**	**V**
6572	**4**	**3**	**2**	**2**	**K**
6589	**4**	**3**	**2**	**2**	**V**

Table 4. Results of SVM, Naïve Bayes and K-NN

Method	Accuracy	Deviation	Kappa
SVM	87.31%	±9.24%	0.814
Naïve Bayes	91.48%	±5.90%	0.876
K-NN	89.39%	±6.86%	0.844

Table 5. Comparison from the Previous Researches

Previous	Method	Kappa
Pantho-Tiantong (2016)	Data Driven	83.40%
Bernard *et al.* (2017)	Data Driven	80.7%
This study	Prior Knowledge	91.48%

Based on the above results (Table 4), the recapitulation of the test results of all methods shows that the Naïve Bayes method obtained the highest accuracy value of 91.48%. This result strengthened deviation value ±5.90, kappa value 0.7, indicates reliable test . The next most accurate results are K-NN, 89.39%, and SVM, with an accuracy value of 87.31%. The detection results using prior knowledge are proven to be better than literature-based and data-driven detection with an accuracy value below 87.31%. This value shows that prior knowledge is proven to provide a representation of students learning styles compared to interaction data from literature-based and data-driven methods. The next step is to use the results of the prior knowledge measurement that the learner has with the attached data.

IV. Conclusion

The method for detecting learning styles with prior knowledge uses assessment data from internal learners. The reliability of this method does not depend on the availability of teaching materials. According to the data, learning style detection using prior knowledge is more accurate than data-driven and literature-based detection. This research needs to be continued by increasing the number of students and should be compared with data from students from different scientific fields.

Acknowledgment

The Ministry of Research, Technology and Higher Education of the Republic of Indonesia funded this study under the Fundamental Research of College Excellence program.

References

[1] V. Pant, S. Bhasin, and S. Jain, ”Self-learning system for personalised E-learning,” in *Int. Conf. Emerg. Trends Comput. Commun. Technol. ICETCCT 2017*, 2018, pp. 1–6. doi: 10.1109/ICETCCT.2017.8280344.

[2] S. Graf and Kinshuk, ”Analysing the behaviour of students in learning styles,” in *Advances in Semantic Media Adaptation and Personalization. Studies in Computational Intelligence, 93*, 2008, pp. 1–23.

[3] S. R. Viola, S. Graf, and Kinshuk, “Detecting learners’ profiles based on the index of learning styles data,” 2002.

[4] S. Graf, S. Viola, and Kinshuk, “Automatic student modelling for detecting learning style preferences in learning management systems,” IADIS Int. Conf. Cogn. Explor. Learn. Digit. Age, no. 1988, pp. 172–179, 2007.

[5] O. El Aissaoui, Y. El Alami El Madani, L. Oughdir, and Y. El Allioui, “A fuzzy classification approach for learning style prediction based on web mining technique in e-learning environments,” Educ. Inf. Technol., vol. 24, no. 3, pp. 1943–1959, 2019, doi: 10.1007/s10639-018-9820-5.

[6] K. Crockett, A. Latham, and N. Whitton, ”On predicting learning styles in conversational intelligent tutoring systems using fuzzy decision trees,” *International Journal of HumanComputer Studies*, vol. 97, pp. 98–115, 2017.

[7] T. Hailikari, “Assessing university students’ prior knowledge: Implications for theory and practice,” tech. rep., University of Helsinki, 2009.

[8] F. J. R. C. Dochy and P. A. Alexander, ”Mapping prior knowledge: A framework for discussion among researchers,” *European Journal of Psychology of Education*, vol. 10, pp. 225–242, Sept. 1995.

[9] D. R. Krathwohl, ”A revision of bloom’s taxonomy: An overview,” *Theory Pract.*, vol. 41, pp. 212–218, Nov. 2002.

[10] O. Pantho and M. Tiantong, ”Using decision tree c4. 5 algorithm to predict VARK learning styles,” *International Journal of the Computer, the Internet and Management*, vol. 24, no. 2, pp. 58–63, 2016.

[11] Fleming, “I’m different; not dumb. modes of presentation (VARK) in the tertiary classroom,” Research and development in higher education.

[12] K. Crockett, A. Latham, D. Mclean, and J. O’Shea, “A fuzzy model for predicting learning styles using behavioral cues in an conversational intelligent tutoring system,” in *2013 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, 2013, pp. 1–8.

[13] J. Bernard, T.-W. Chang, E. Popescu, and S. Graf, ”Learning style identifier: Improving the precision of learning style identification through computational intelligence algorithms,” *Expert Syst. Appl.*, vol. 75, pp. 94–108, June 2017.

[14] M. Abdullah, W. H. Daffa, R. M. Bashmail, M. Alzahrani, and M. Sadik, ”The impact of learning styles on learner’s performance in e-learning environment,” *J. Adv. Res. Med.*, vol. 6, no. 9, pp. 24–31, 2015.

[15] S. V. Kolekar, R. M. Pai, and M. M. M. Pai, ”Prediction of learner’s profile based on learning styles in adaptive e-learning system,” *International Journal of Emerging Technologies in Learning*, vol. 12, no. 6, pp. 31–51, 2017.

[16] P. Garcia, A. Amandi, S. Schiaffino, and M. Campo, ”Using bayesian networks to detect students’ learning styles in a webbased education system,” in *7º Simposio Argentino de Inteligencia Artificial - ASAI2005*, Aug. 29–30, Rosario, 2005, pp. 115–126.

[17] E. Ozpolat and G. B. Akar, “Automatic detection of learning styles for an e-learning system,” *Comput. Educ.*, vol. 53, pp. 355–367, 2009.

[18] M. D. Calvo-Flores, E. G. Galindo, M. C. P. Jimenez, and O. P. Pineiro, ”Predicting students’ marks from moodle logs using neural network models,” in *Current Developments in Technology-Assisted Education*, 2006.

[19] X. Wu, V. Kumar, J. Ross Quinlan, J. Ghosh, Q. Yang, H. Motoda, G. J. McLachlan, A. Ng, B. Liu, P. S. Yu, Z-H. Zhou, M. Steinbach, D. J. Hand, and D. Steinberg, ”Top 10 algorithms in data mining,” *Knowl. Inf. Syst.*, vol. 14, pp. 1–37, 2008.

[20] Y.-C. Chang, W.-Y. Kao, C.-P. Chu, and C.-H. Chiu, ”A learning style classification mechanism for e-learning,” *Comput. Educ.*, vol. 53, pp. 273–285, 2009.

[21] J. Feldman, A. Monteserin, and A. Amandi, ”Detecting students’ perception style by using games,” *Comput. Educ.*, vol. 71, pp. 14–22, 2014.

[22] Fletcher, ”Support vector machines explained,” Tutorial paper, 2009.

[23] K. Madadipouya, ”A new decision tree method for data mining in medicine,” *Adv. Comput. Intell. Int. J. (ACII)*, vol. 2, no. 3, pp. 31–37, 2015.

[24] M. S. Hasibuan, L. E. Nugroho, and P. I. Santosa, ”Model detecting learning styles with artificial neural network,” *J. Technol. Sci. Educ.*, vol. 9, no. 1, 2019.

[25] T. Hailikari, N. Katajavuori, and S. Lindblom-Ylanne, ”The relevance of prior knowledge in learning and instructional design,” *Am. J. Pharm. Educ.*, vol. 72, no. 5, 113, 2008.