Identification of Learning Characteristics Pattern of Engineering Students using Clustering Techniques

Aisyah Larasati[1], Apif Miftahul Hajji[2] and Anik Nur Handayani[3]
[1] *Department of Industrial Engineering, Faculty of Engineering, Universitas Negeri Malang, Malang, Indonesia*
[2] *Department of Civil Engineering, Faculty of Engineering, Universitas Negeri Malang, Malang, Indonesia*
[3] *Department of Electrical Engineering, Faculty of Engineering, Universitas Negeri Malang, Malang, Indonesia*

Keywords: Learning Characteristics, Engineering Students, Data Mining Clustering Technique.

Abstract: Everyone has their own characteristic way of thinking that make them to have different ways to act. These characteristics also affect their behaviour in daily life, including their learning characteristics. This study aims to identify the learning characteristics pattern of engineering students using data mining clustering technique. This study uses questionnaire to collect data. The total number of students fill out the questionnaire are 2,934. After data preparation steps, only 1,914 responses (65.23% usable rate) are complete and can be used for further analysis. To identify the learning characteristics pattern, this study uses data mining clustering technique. The clustering techniques used in this study are K-means cluster, Kohonen cluster analysis, and two step cluster analysis. The results show that all three cluster techniques used in this study identify the frequency of a respondent does an independent study by solving practice exercise after learning a new material in the class, the frequency of a respondent studies the material he learnt after attending a class and the frequency of a respondent discusses the learning material are the top three important variables to differentiate each cluster.

1 INTRODUCTION

Everyone has their own characteristic way of thinking. These characteristics make them to have different ways to solve a problem and a special process to identify an issue. Of course, this also affect their behaviour in daily life, including their learning characteristic. How they react to many issues in daily or what they choose to make some decisions on learning process. It creates human to become a complex creature. The human behaviour itself is a human process to act and interact with each other. It always becomes a complex process that hard to understand, considering human behaviour depends on many life factors. Human behaviour can determine people to take some decision and create a habit on daily basis. According to Icek Ajzen theory, human behaviour guided by three consideration: beliefs about the result of their behaviour and the evaluation of their result (behavioural beliefs), beliefs about other people expectations and motivate to do it (normative beliefs), and beliefs about factors that can inhibit or facilitate the accepted behaviour and the impact of it (control beliefs) (Ajzen, 1985). What people experience in daily life will create their behaviour pattern. People will show their behaviour with their actions or how they interact with each other.

The world of education is also changing due to the global internet phenomenon. Internet gives people much accessible media for learning source. A large number of accessible media changes college student behaviour to learn. Basically, students have their own reason to choose how they learn, what their learning styles, which media will be used, etc. Students media usage behaviour is strongly influenced by three factors: sociability, utility and reciprocity (Zawacki-Richter *et al.*, 2015). Sociability can be reviewed by their interaction with each other which lead to selected media. The utility is processing to get the best result with by using accessible media with maximal effort. The last, reciprocity is how they use accessible media to improve their cognitive skill by reading, being critical, and understanding. Of course, not all of them choose internet for media to learn. A few college students still feel better to understand what they learn with a book or any non-internet media.

Larasati, A., Hajji, A. and Handayani, A.
Identification of Learning Characteristics Pattern of Engineering Students using Clustering Techniques.
DOI: 10.5220/0008411002740278
In *Proceedings of the 2nd International Conference on Learning Innovation (ICLI 2018)*, pages 274-278
ISBN: 978-989-758-391-9

The difference in learning characteristics is influenced by individual characteristics as well. In addition to behaviour, personal characteristics play a major role in the learning characteristics. Personal characteristics show special behaviour in each individual. Various studies have revealed the importance of understanding student characteristics on the effectiveness of the learning process (Kauffman, 2015; VanSickle *et al.*, 2015; Apple, Duncan and Ellis, 2016). However, there is still not much literature that explore student characteristics by applying data mining technique. According to (Sin and Muthu, 2015), data mining techniques can be used to improve academic quality, including predicting student performance in learning, data visualization, detecting student failures in learning and even investigating student behaviour in learning. Therefore, data mining techniques are also potential to be used to segment student characteristics in order to understand their learning behaviour. This study aims to identify the learning characteristics pattern of engineering students using data mining clustering technique. The cluster resulted from this research can be used to figure out the existing differences among cluster and provide faculty members with some insight of their student characteristics.

2 METHOD

This study uses an online questionnaire to collect data. The online questionnaire is administered via Universitas Negeri Malang Academic Information System (SIAKAD) in April - May 2018. The target respondents are all registered students in Faculty Engineering at Universitas Negeri Malang, which are approximately 5,300 students. Among of those registered students, only 2,934 students fill out the questionnaire (55.34% participation rates). The data mining clustering model is built following the SEMMA procedure, which are Sample, Explore, Modify, Model, and Assess.

The first step, sample is conducted by determining the target object of the study, which are all registered students in Universitas Negeri Malang. The explore step aims to understand the nature of collected data, which is performed by plotting the collected data. The modify step includes data preparation and data transformation when needed. Data preparation steps include data cleaning and data imputation. Data cleaning aims to delete uncompleted responses and outlier responses. Data imputation is performed to impute missing responses with the mode or mean responses. After data preparation steps, only 1,914 responses (65.23% usable rate) are complete and can be used for further analysis. To identify the learning characteristics pattern, this study uses three clustering data mining technique. The clustering models built in this study are K-means cluster, Kohonen cluster analysis, and Two step cluster analysis. The last step is to determine how to assess the model performance (accuracy). Regarding the model accuracy, this study use Silhouette index as suggested by (Pereda and Estrada, 2018). Silhouette measures distance of an element to its own cluster (cohesion) and compares it to other clusters (separation). The higher value indicates that the element is well matched to its own cluster and poorly matched to other defined clusters.

The questionnaire that is used to collect data contains 4 questions about respondent's profile (gender, level of study period, study program, and GPA). In addition, the questionnaire contains 18 closed-ended question that ask about the learning characteristic of respondent. Briefly, the item list of the questions in the questionnaire is shown in Table 1.

Table 1: Item list in the questionnaire.

Variable	Indicators	Number of Items
Learner Characteristics	Personal Profile	4 open-ended questions
	Learning preparation	1, 2
	Initial understanding	3, 4, 5, 6
	Characteristics of discussion activity	7, 8, 9, 10, 11
	Understanding during learning process	12, 13, 15
	Characteristics of independent study	14, 16, 17
	Learning styles	18

Table 2: Descriptive statistics of the respondent's profile.

Department	Count (respondent)	Percentage (%)
Civil Engineering	423	22.10%
Electrical Engineering	671	35.06%
Mechanical Engineering	500	26.12%
Industrial Technology	320	16.72%

The descriptive statistic of the personal respondents profile is shown in Table 2. In addition to the department of the respondents, the profile also show that the respondents are students in the first years up to the seventh years of study in the faculty of engineering who has GPA from 2.00 – 4.00.

3 RESULT AND DISCUSSION

The results of the classification using data mining techniques as shown in Table 3 indicate that K-means technique results in the highest number of cluster, meanwhile Two Steps techniques results in the lowest number of cluster. However, the Kohonen technique results in the highest range of the size of cluster. Based on the range of the cluster size, it seems that Kohonen technique works best to classify students learning characteristics. This results implies that it is better to classify the learning characteristics into five clusters: cluster 1 (20.34%), cluster 2 (30.46%), cluster 3 (6%), cluster 4 (14.54%), and cluster 5 (28.66%). The largest cluster, cluster 2, is dominated by students who are moderate frequently do an independent study. On the other hand, cluster 4, is dominated by student who are always do an independent study. While in other cluster, the students are rarely do an independent study.

The model accuracy is measured based on the silhouette cohesion and separation index. The silhouette indices of the three clustering models in this study imply that the best clustering method is the two steps model since this model results in the highest Silhouette index. On the other hand, K-means cluster has the lowest Silhouette index. This indices indicate that the Two steps model has the best cohesion and separation ability compared to K-means and Kohonen models.

Table 3: Comparison of the classification results using different techniques

Model	Number of Clusters	Smallest Clusters (%)	Largest Clusters	Silhouette Index
K-means	12	4	14	0.21
Kohonen	5	6	30	0.63
Two steps	2	46	53	0.82

Detail results shown at Figure 1, Figure 2, and Figure 3 indicate that all three cluster techniques used in this study identify var 14 (how frequent a respondent does an independent study by solving practice exercise after learning a new material in the class), var 15 (how frequent a respondent studies the material he learnt after attending a class) and var 11 (how frequent a respondent discusses the learning material) are the top three important variables to differentiate each cluster.

An independent study, either by doing some exercise or studying learning material, is a form of an active learning activity. An independent study improve student performance in undergraduate science, technology, engineering, and mathematics (STEM) course (Freeman *et al.*, 2014). Students' intention to do independent study may vary across all department in Faculty of Engineering. Thus, solving practise exercise and repeating studying learning material become important variable to cluster students learning characteristics.

Discussion in learning process is an activity that requires a student to express his/her thought to other and gain feedback on it. Discussion activity include two main process, an external interaction and an internal process (Illeris, 2009). The external interaction means a student have to interact to his/her teacher/peer, or surrounding environment. The internal process related to his/her psychological ability to elaborate and acquire information into his/her mind. Discussion intention might be significantly differently among the respondents, thus it become one of the top three important variables to cluster them. This result is also in line with other study that find relationship between frequency of discussion with tutor and learning outcomes. Learning characteristics, which include self-efficacy and inside knowledge affects the structure of the discussion (Mitchell *et al.*, 2013).

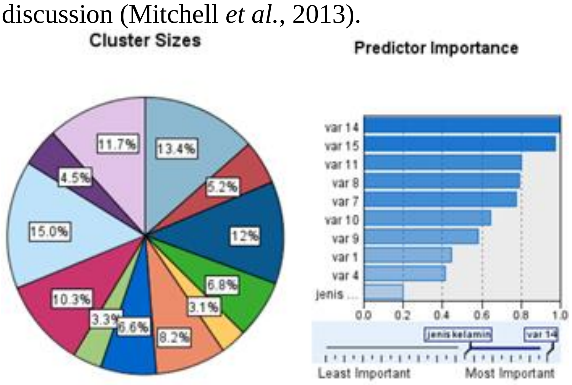

Figure 1: Cluster size and predictor importance based on k-means technique.

Figure 2: Cluster size and predictor importance based on kohonen technique.

In addition, the three cluster techniques used in this study do not include var 2 (a respondent motive to do learning preparation before attending a class), var 3 (a respondent learning style while developing early understanding of the learning material), var 5 (the reason behind a respondent behaviour to develop early understanding of the learning material) and var 6 (a respondent learning style to prepare her/himself before attending a class) as important variables to cluster the respondents. It means that there is no significant difference among respondents based on these four variables. This results implies that the students' motive, learning styles, intention to develop early understanding is vary across all department in Faculty of Engineering. There is no specific pattern that may be used to classify the respondent.

Figure 1 shows the least important factor to cluster the respondent using K-Means technique is gender. This result implies that female and male are only slightly different in their learning characteristics according to K-Means cluster. This result is in accordance with (Subramanian, 2018) findings that indicate no differences between male and female students in their learning styles. Figure 2 indicates that the least factor to cluster the respondent using Kohonen technique is var 13 (the activity that a respondent do when he/she does not understand a learning material). This finding implies that there is only slightly different on the student activity that are performed by the students when they don't understand the learning material. The individual responses show that most students (859 students) tend to ask their friend instead of doing the other available options. In addition, the Two Step technique identify var 17 (a place where a respondent usually does an independent study) as the least important variable. This result implies that students' preferences on choosing the place to study is only slightly different among cluster. It is supported by the individual responses that reveal the study area provided in the campus is the most favourite place to study.

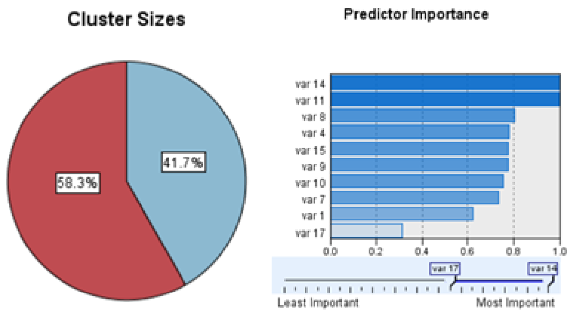

Figure 3: Cluster size and predictor importance based on two steps technique.

4 CONCLUSIONS

Based on Silhouette index, the best model to cluster student learning characteristics is the Two Step Model. The top three students learning characteristics that are important to differentiate one cluster to the other cluster are: 1) the frequency of a student does an independent study by solving practice exercise after learning a new material in the class; 2) the frequency of a student studies the material he learnt after attending a class; and 3) the frequency of a student discusses the learning material. On the other hand, the other learning characteristics that are only slightly different between one cluster to the other clusters are: 1) a student motive to do learning preparation before attending a class; 2) a student learning style while developing early understanding of the learning material, the reason behind a student behaviour to develop early understanding of the learning material; and 3) a student learning style to prepare her/himself before attending a class.

ACKNOWLEDGEMENTS

Acknowledgments to Universitas Negeri Malang (UM) which has funded this research through Islamic Development Bank (IsDB) - UM Research Grant.

REFERENCES

Ajzen, I. (1985) 'From Intentions to Actions: A Theory of Planned Behavior', in Kuhl, J. and Beckmann, J. (eds) *Action Control: From Cognition to Behavior*. Berlin, Heidelberg: Springer, pp. 11–39.

Apple, D., Duncan, W. and Ellis, W. (2016) 'Key Learner Characteristics for Academic Success', *International Journal of Process Education*, 8(2).

Freeman, S. *et al.* (2014) 'Active learning increases student performance in science, engineering, and mathematics', *Proceedings of the National Academy of Sciences*, 111(23), pp. 8410–8415.

Illeris, K. (2009) 'A comprehensive understanding of human learning', in Illeris, K. (ed.) *Contemporary Theories of Learning: Learning Theorists ... In Their Own Words*. New York: Routledge, pp. 1–14.

Kauffman, H. (2015) 'A review of predictive factors of student success in and satisfaction with online learning', *Research in Learning Technology*, 23.

Mitchell, C. M. *et al.* (2013) 'Learner characteristics and dialogue: recognising effective and student-adaptive tutorial strategies', *International Journal of Learning Technology*, 8(4), pp. 382–403.

Pereda, M. and Estrada, E. (2018) 'Machine Learning Analysis of Complex Networks in Hyperspherical Space', *eprint arXiv:1804.05960*.

Sin, K. and Muthu, L. (2015) 'Application of big data in education data mining and learning analytics--a literature review', *ICTACT Journal on soft computing*, 5(4).

Subramanian, S. (2018) 'Learning styles and achievement in english among higher secondary school students–gender wise analysis', *Indian Journal Of Applied Research*, 8(8).

VanSickle, J. L. *et al.* (2015) 'An Investigation of Student Performance, Student Satisfaction, and Learner Characteristics in Online Versus Face-to-Face Classes--RESEARCH', *Kentucky Journal of Excellence in College Teaching and Learning*, 13(1), p. 3.

Zawacki-Richter, O. *et al.* (2015) 'Student media usage patterns and non-traditional learning in higher education', *The International Review of Research in Open and Distributed Learning*, 16(2), pp. 136–170.