Finding Contributing Factors of Students' Academic Achievement Using Quantitative and Qualitative Analyses-Based Information Extraction

https://doi.org/10.3991/ijet.v17i16.31945

Ariana Yunita[1,2], Harry B. Santoso[1(✉)], Zainal A. Hasibuan[3]
[1] Universitas Indonesia, Depok, Indonesia
[2] Universitas Pertamina, Jakarta, Indonesia
[3] Universitas Dian Nuswantoro, Semarang, Indonesia
harrybs@cs.ui.ac.id

Abstract—Big data learning analytics is still in its infancy and has been developed on several campuses worldwide. Ideally, all students' profiles should be described and embraced to optimize the development of any proposed system related to big data learning analytics. This paper aims to extract information related to factors contributing to students' academic achievement using quantitative and qualitative approach, in which co-occurrence analysis were applied for quantitative approach and facet analysis for the qualitative approach. For data collection, Kitchenham's technique were used to select and filter the literature, at the first iteration, 1,167 papers were found. After applying inclusion and exclusion criteria, 101 articles were processed for text mining. Titles and abstracts were analyzed using a text-mining tool, and then resulted clusters of words. Afterwards, clusters of words were labeled using facet analysis. This study results in eight interrelated clusters of academic achievement factors: demography, internal consistency, technology, student course engagement, activity in a classroom, educational system, socio-culture, and personality. Several insights into each cluster will be described and might be beneficial for researchers in learning analytics.

Keywords—text mining, big data learning analytics, student's success, facet, co-occurrence analysis

1 Introduction

In recent years, the research of big data analytics has been gaining significance in every field. It is due to the growth of data and the increase of computing power that has enabled researchers to embrace advanced technology in processing the data that meets the criteria of 5V's (Variety, Veracity, Volume, Velocity, and Value)[1]. Big data is essential for educational institutions [2] in which the application is for analyzing learners' behavior [3]–[7] to gain students' profiles, improving students' retention [8], [9] and evaluating student's feedback [10]–[12]. However, since learning is a complex process [13] with various factors affecting students and many stakeholders in the learning process, it might be beneficial to revisit and define what type of features should be included in the conceptual framework for predicting student's academic achievement.

Research related to predicting student's academic achievement has been a concern since five decades ago. In 1940, a study reviewed several factors on academic achievement and claimed that the most contributing factors affecting student's grades were intelligence and academic motivation [14]. These findings had not been changed until the 1990s. Several researchers agreed that those two factors play a significant role in academic achievement [15]—other research-related academic achievements with psychological aspects, such as personality and learning styles. In line with the development of the internet and technology, learning media also vary. Up until now, technology-enhanced learning, such as learning management systems, virtual learning, Massive Open Online Course (MOOCs), has emerged to support the learning process in which each learning environment contains several success factors [16].

A study found factors contributing to academic achievement in online learning [17] and resulted in interrelated entities to support student retention. Instead of focusing on online education, this study revisits and finds factors related to student retention in online learning and broaden the context of learning in general. This study uses a different approach from the previous studies and results in various student factors. While the previous research using Kitchenham's Systematic Literature Review (SLR), this study applies the strategy of seeking knowledge using literature-based discovery [18], then analyzing using quantitative and qualitative approach. This study aims to describe factors affecting students' academic achievement using a co-occurrence analysis and using facet analysis to propose a conceptual model related to big data learning analytics.

The organization of this paper is as follows. Section 2 explains the literature review; Section 3 describes the research methodology; Section 4 discusses the results; Section 5 describes the limitations of this study, concludes and recommends for future study.

2 Literature review

The next subsections provide the theoretical foundations and literature review that is underlying in this study.

2.1 Big data and learning analytics

Learning Analytics (LA) is the research area that is proposed by a community called SoLAR (Society for Learning Analytics Research). LA's main idea is using educational data to improve the learning process [19] by analyzing activity data taken from the learner and all agents involved in a learning process. In general, LA is used to optimize the learning process. Several benefits from LA are improving curriculum, optimizing student learning outcomes, identifying learning habits and learning processes, supporting personalized learning, improving instructor performance, analyzing post-educational work, and contributing to education [20].

Current research review regarding big data usage for LA recommend that future research will focus on developing automated systems that can acquire, analyze, and aggregate large amounts of data, then produce descriptive, predictive, and prescriptive analysis for learning analytics purposes [21]. There are four objectives of current research related to big data learning analytics: to improve learning process, to analyse learning behavior, to improve student retention and to evaluate learning process. In short, big data has been considered to be embraced in learning analytics [21], and transdisciplinary approaches should be carried out [13], [22].

2.2 Text mining

Text Mining is an interdisciplinary field that draws on information retrieval, data mining, machine learning, statistics, and computational linguistics [23]. A large portion of the data, such as news articles, technical papers, books, digital libraries, email messages, blogs, and web pages, is stored as text. Documents need to be ranked for more efficient retrieval over extensive collections in many text mining applications, particularly information retrieval (IR). Documents are represented as vectors with a numerical value assigned to each word in order to describe the significance of a word in a text.

In LA, text mining has been used for various applications, such as opinion mining from students' feedbacks [10]–[12]. A study evaluated the course using data generated from the students' survey, then classified the feedback as negative or positive [10]. Furthermore, another study also implemented text mining to assess students' satisfaction, not only for sentiment analysis but also for monitoring learners' satisfaction using big data approaches [11]. In the same way, [12] proposed a big data framework to evaluate students' opinions. The differences with [11], [12] also tried to predict student's emotions. However, the data used by Jena [12] is still unclear due to the disclosure agreement. Another study [24] also applied text mining and proposed a recommender system for an intelligent chatbot. Another study applied text mining in a different context, Rahmah et al. [18] conducted literature-based-discovery that combined literature review and text mining to find Technology-Enhanced Learning components.

Previous studies show that research regarding big data learning analytics attempted to embrace all possible data [21]. This study adopt the research method proposed by Rahmah et al. [18] that combine literature review and text mining, but we use a different text mining tool. Section 3 shows the research methodology for extracting information of factors contributing to students' academic performance.

3 Methodology

Our main objective of this study is to find factors that contribute to student's academic achievement. For data selection, this study used the technique to select papers using Kitchenham's technique. Figure 1 shows the step-by-step of doing this research. Mainly there are four phases: planning, data collection, data analysis, and interpretation. Firstly, in the planning phase, we emphasize the research problem and research objective. A further step is collecting published studies using Kitchenham's approach to select papers and mining the title and abstract. The third step is analyzing the results, and the last is interpreting the results.

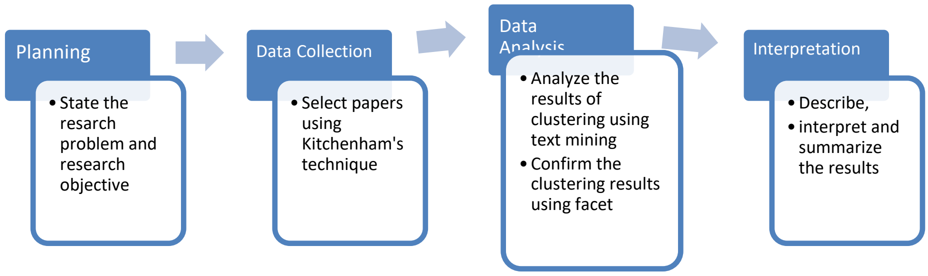

Fig. 1. Research phases

3.1 Data collection

A set of documents that contain factors affecting students' academic achievement were selected and filtered using Kitchenham's guideline [25]. Firstly, the searching strategy and keywords were described. Several words are derived from the research question, and several synonyms of the terms are determined to be used in the search strategy. The first term that might be included is "predict", because students' factors were commonly used to predict students' academic performance [26]–[28]. After several iterations, it can be detected that several studies use the word predicting, prediction, forecast, detect, detecting, detection, asses, assessing, assessment, analyze or analyze to replace the term predict. Another term is academic achievement, which is very similar to academic performance where several studies use it interchangeably. Other synonyms for those two terms are retention, dropout, and attrition. Academic or student are commonly at the beginning of those terms, such as academic achievement or student achievement.

From the explanation above, keywords used for this review are: (predict* or forecast or detect* or assess* or analy*) AND (academic or student) and (performance or achievement or attrition or retention or dropout) and (model or system or framework or technique or algorithm) and (internal or external) and factor*. This study's database is Scopus because it contains an extensive abstract and citation database of peer-reviewed literature. Google Scholar and Web of Science were not selected in this study because of several reasons. First, Google Scholar includes peer-reviewed literature and all manuscripts without reviewed by other researchers, such as academic articles, proceeding papers, and journal papers. Second, most of Web of Science's publications are not peer-reviewed, while Web of Science is mostly overlapping with Google Scholar [29].

A further step is defining inclusion and exclusion criteria. The criteria are: papers were published from 2015 until 2020, the publications must be in English, only articles were published in proceedings or journals. The detailed steps of selecting documents are explained in Figure 2 below.

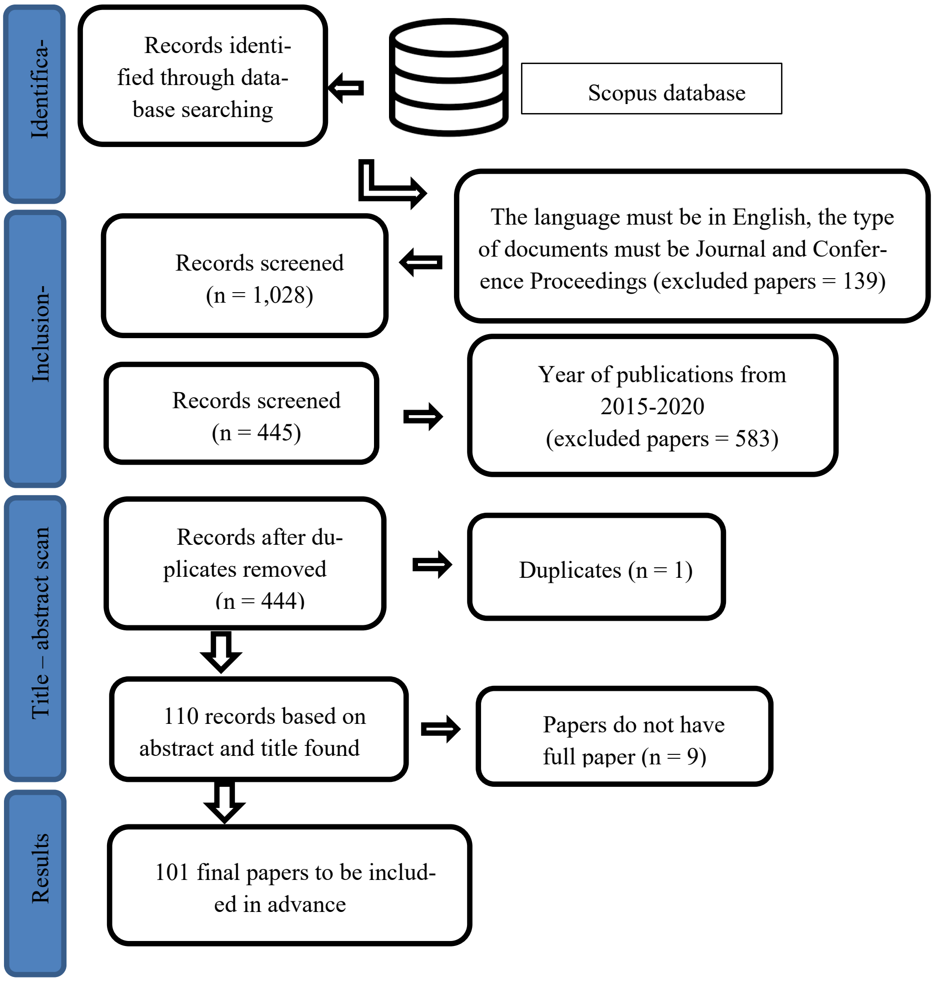

Fig. 2. Selection process of potential papers

Searching with the specific keywords was conducted on 14 February 2020. Firstly, the year of publication was not set, and the search resulted in 1,167 documents. Publications for this keyword year-by-year and distribution of country can be seen in Figure 3. Figure 3 shows that research regarding predicting students' academic achievement has grown from 1970 until now.

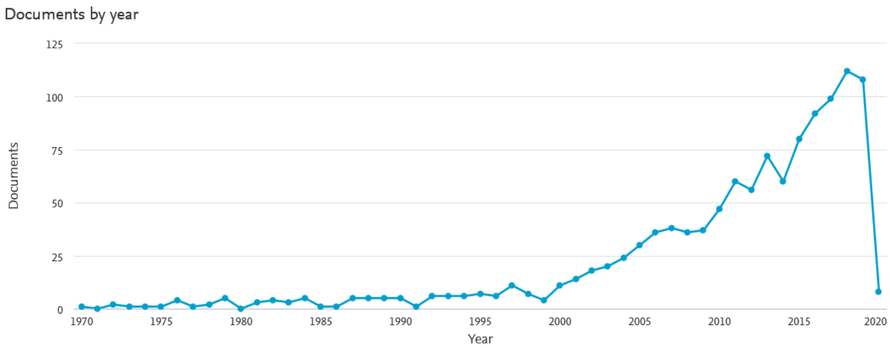

Fig. 3. Research trend year-by-year

For the next step, the results were limited with two criteria: language must be in English, the type of documents must be Journal and Conference Proceedings. Due to the process, it resulted 1,028 documents. The year of publications were limited from 2015 to 2020, and it resulted 445 documents. Due to similar publications, it resulted in 444 documents. We scanned the title and abstracts to find a set of publications that seems to answer the research question. From this step, 109 documents were collected. Furthermore, papers that have no full text were disregarded. Finally, 101 documents were processed using text mining. The process of text mining is explained in the next section.

3.2 Data analysis using co-occurrence analysis and facet analysis

For the quantitative approach, the analysis was carried out with the help of VOSviewer clustering functions, which aid in calculating similarities between key terms based on their association strength and a weighted sum of squared distance. Even though Vosviewer is developed for bibliometrics analysis [30], this application also can be used for text mining [31] to detect the occurrences of a term from abstract and title. Several parameters were set in Vosviewer, such as choosing the minimum number of term's occurrences and the number of terms to be selected. The details of the parameters selected are shown in Table 1.

Table 1. Vosviewer Parameters Chosen for This Study

Parameters	Selection for This Study
Fields from which terms were extracted	Title and abstract
Full counting/binary counting	Full counting
Minimum number of occurences of a term	3 (of the 3,334 terms, 475 terms meet the threshold)
Number of terms to be selected	90%
Method	Fractionalize
Fields from which terms were extracted	Title and abstract
Full counting/binary counting	Full counting

Next, we deselected irrelevant words manually, for example, the name of publications: Taylor & Francis Group, Springer, and Elsevier. We also disregarded several terms about research methodology, such as validity, design, approach, control group, methodology, framework, correlation, hypothesis, sample, evidence, mean, regression, questionnaire, structural equation, and survey. Several terms common in the abstract were also ignored, such as introduction, method, purpose, aim, analysis, limitation, and conclusion.

After deselecting irrelevant terms, terms were clustered using a unified framework for mapping and clustering in Vosviewer [31]. The results from Vosviewer shows several clusters, where each cluster should represent each topic, and some insights could be gained.

The next process is analysing using Facet in which the analysis summarized as follows: It provides a rational, scientific methodology for the construction of systems; it allows for the full and precise description of objects with significant structural complexity and multi-dimensional semantic composition; and it provides a flexible syntactical apparatus for the combination and ordering of concepts where this is required [32], [33]. This approach's strength lies in its logical principles and the way it provides structures in knowledge organization systems (KOS). The main flaws are (1) its lack of empirical foundation and (2) its speculative ordering of knowledge without reference to the development or influence of theories and socio-historical studies [34].

4 Results and discussion

The document selection process and text mining results in the network map, as shown in Figure 4, and a list of words in each cluster, as shown in the Appendix. Using quantitative approach, it resulted in five clusters, and we carefully analyzed and labeled using Facet [34] as "Cluster Demography-Internal Consistency", "Technology-Course Engagement", "Activity in a Classroom- Educational system- Intrinsic Motivation", "Socio-culture", and "Personality". Each cluster is analyzed and explained below.

4.1 Demography-internal consistency

The first cluster might be labeled as structure, measure, age, difference, association, or internal consistency, as the node has the most considerable edge, in other words. However, when we traced back to the literature, internal consistency might be the most appropriate to represent this cluster. Internal consistency is also known as self-regulation. As can be seen in Figure 4 that the term internal consistency relates to adaptation, and adaptation relates to self-regulation. Self-regulation is considered necessary in academic success [24, 28, 35], not only for learning that has characteristics to change stimulus-response such as playing music [28] and cognitive learning process. However, it is essential to note here that when traced back to literature containing "internal consistency", it also refers to internal consistency related to validation when conducting a survey using a specific questionnaire. Regarding Self-

Regulated Learning (SRL), it is interesting to note here that since traditional learning shifted to online learning, research regarding SRL has also grown to online learning, since SRL can be traced in the context of formal education online learning [35].

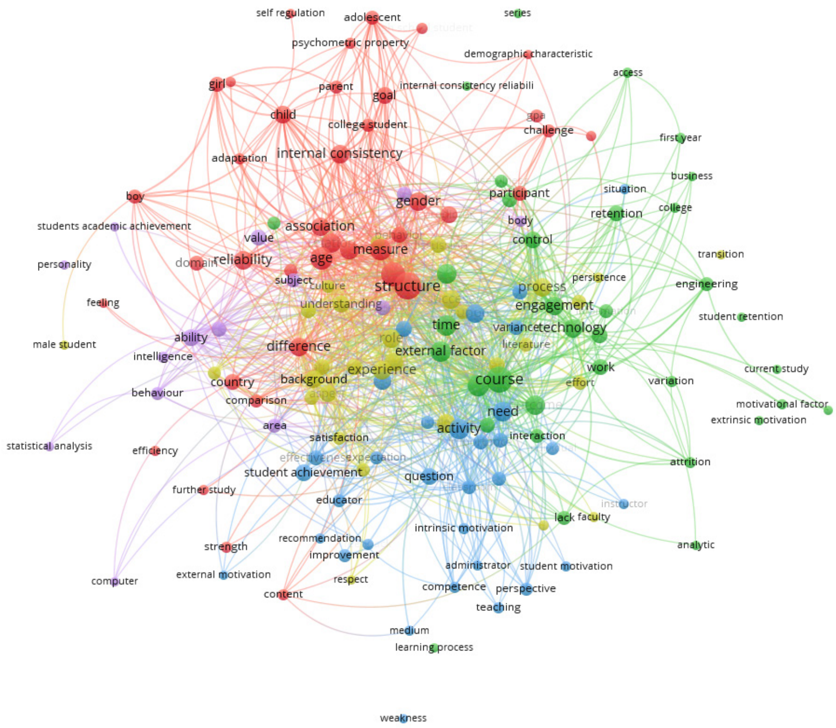

Fig. 4. Network map of students' factors

Another term that might reflect this cluster is demographic characteristics because the word "internal consistency" relates to several terms, such as gender, child, adolescent, age, adaptation, goal, and psychometric property. It can be argued that demography characteristics, which consist of several terms related to demography, dominate this cluster. Gender is a predictor of student success as boys tend to have lower academic achievement [36]. Age also essential to be a predictor as it might differentiate behavior, time management, working activity between adolescents and kids.

Due to the mixed topics in this cluster, this cluster can be divided into Demography and Internal Consistency. Even though other terms seem to be indifferent, such as behavior, family, evaluation, college, and strength, the demography characteristics and internal consistency are still dominant. It can be seen in Figure 4 as the terms internal consistency seems to be a relatively larger node than others.

4.2 Technology-student course engagement

The second cluster contains several terms related to engagement, course, technology, and work. Regarding technology, the previous study pointed out that it might negatively impact students, for example, internet addiction [37]–[39]. Nevertheless, technology has a more significant number of benefits than the consequences. Technology provides an environment for educators and students, so learning and teaching have no boundaries in time and place. All applications that use technology to facilitate learning, called Technology-Enhanced Learning (TEL). The implementation of TEL is a mobile learning management system that influences online students' academic achievement [40], e-learning, multimedia learning environment [41], MOOC [42]. However, several issues might arise with technology-enhanced learning, such as usability factors in learning media [43], an external motivation that affects learners [44], and one size fits all dashboards [45].

Regarding student course engagement, the term related to student course engagement is 'working'. A study states that students with part-time jobs more than 20 hours per week are less likely to complete degrees [46]. Another term related to course engagement is self-efficacy, which is defined as an individual belief in his ability to succeed in doing something and consequently affect individual behavior, persistence, intention, commitment, and effort [47][48]. Another study agrees that intention and commitment might be the two issues that affect academic achievement [49].

4.3 Activity in a classroom-educational system

This cluster relates to the educational system as several words in this cluster are activity, class, classroom, educator, and teaching. From the literature, it can be concluded that the educational system affects student's performance, including type of school [50], teaching strategies [51], assessment [52], and teacher [53]. Those factors are interrelated, such as administrator to teaching, teaching to activity, task to classroom, and task to engagement. A classroom as a learning environment also should be a priority for students' success. A study shows that planting trees in school backyards, especially tree cover and tree diversity, affect students' academic performance[54].

Other options for labeling this cluster are intrinsic motivation cluster. From the list of terms in the Appendix, intrinsic motivation is clustered in the third, while extrinsic motivation in the second cluster. Several terms that highly relate to intrinsic motivation are mindset [55] and academic self-concept [56], [57]. However, according to literature [58], motivation is one of the aspects of SRL, therefore motivation was not a label in this cluster.

4.4 Socio-culture

As shown in Appendix, this cluster contains thirty terms, mostly related to socio-cultural factors, such as background, culture, experience, social, and attitude. Socio-cultural factors play a significant role in academic achievement motivation as Western students differ from Asia in terms of academic self-concept. Hence, it may affect motivation, and consequently, motivation influences academic achievement [59]. Another study finds that low violence at school and family socioeconomic status relate to academic achievement [60]. In line with that, the social environment also influences students. A study finds that students who are not significantly affected by the institution's social environment, such as university study after working for some time, commuter, part-time students relate to their performance [61]. In contrast, [62] point out that parents and peers were the two most significant social support for undergraduate students within the age range of 18-24.

This cluster shows that social relations are essential to academic achievement. However, the term family is not included in this cluster, but in the first cluster. Family support, especially in the first semester, is considerably influential in student retention [63].

4.5 Intelligence, behavior, ability, personality, interest - personal factors

This cluster contains fourteen terms that most relate to students' themselves, such as intelligence, behavior, ability, personality, interest, and value. First, term intelligence, when this term was traced into the literature, this term related with emotional intelligence [64], artificial intelligence [65]-this terms might not represent student's factors because it refers to a specific research topic, intrapersonal intelligence [63], verbal intelligence [66], and social intelligence [67].

Another term is "ability", which is the subset of several terms, such as cognitive ability [41] and emotion regulation ability [68]. Furthermore, the term is interest; this term reveals that students' interest relates to motivation and significantly affects academic performance. Other terms, such as behavior related to planned behavior that mostly affected by intention. Positive or negative motives are created by the perception of how other people will view similar actions (external forces) [69].

Based on the analysis, this cluster seems to be suitable to be labeled as a personality. Personal factors, especially the interests of students [57], are essential predictors for academic achievement. Also, previous academic achievement is a significant predictor in Medical students, even though it might not be generalized to every student [70].

From the results of quantitative approach, each cluster's factors in each cluster interrelated mean each cluster's cluster factor in a cluster might relate to other factors in the same cluster and a different cluster as can be seen in Figure 5. Afterwards, we analysed and split carefully. Then, it resulted eight clusters: demography, internal consistency, technology, student course engagement, activity in a classroom, educational system, socio-culture, and personality.

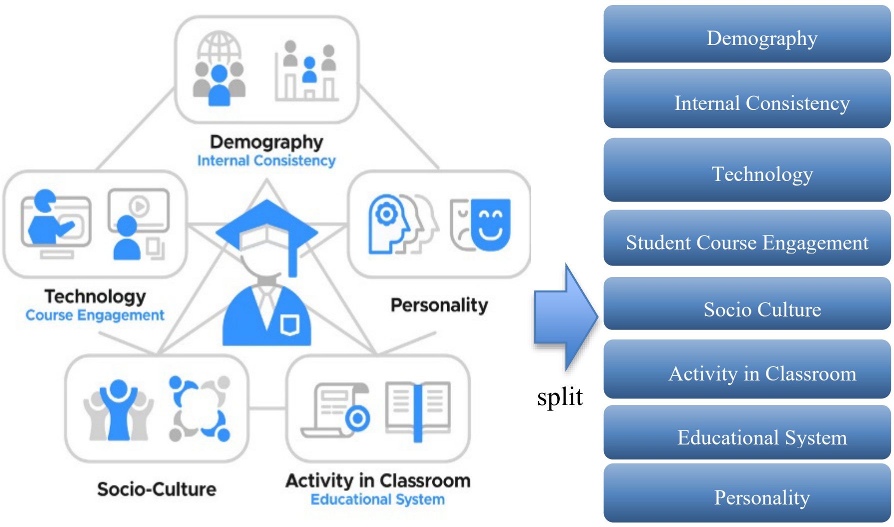

Fig. 5. Cluster of students' factors

5 Limitations and conclusion

This study has several limitations regarding the data selected and the parameter in the quantitative tools. First, related to the data selected in this study were only papers within 5 years. There should be more than five years to get another result. Keywords and the inclusion and exclusion also could be different. The second limitation is related to parameter chosen in Vosviewer. Different parameter chosen in Vosviewer might lead to different clusters then it might result biases while retrieving literature on automatic search. Even though this study has several limitations, as explained above, this result might still benefit all stakeholders in the learning process as the implication might lighten students, teachers, policymakers, and researchers interested in education.

First, for students, they might consider that their academic achievements are not only affected by themselves. Neither their intrinsic motivation nor their personality is a single factor that affects their academic achievement. Various factors, including assessment process, teaching process, student engagement in class, the technology used for learning, age, family support, and other factors, are essential.

Second, for teachers, they might consider that even though they have provided better teaching, the result will not always be better due to many factors that affect students.

Third, for policymakers, since many factors of students' factors, they also might consider supporting the education and learning process in various aspects, such as teacher, technology, and providing a scholarship to help students gain the better achievement.

Fourth, for researchers in learning analytics, especially related to student retention, they might portray students' factors as large as possible.

To conclude, this study attempted to extract information using dual approach, which are co-occurrence analyses and facet analysis. In this study, the dual approach produces eight clusters of students' factors: Demography, Internal Consistency, Technology, Students Course Engagement, Socio-culture, Activity in Classroom, Educational System, and Personality. This study is being a foundation for further study to collect data from multiple sources and build a model for adopting big data in learning analytics. This study might also help researchers in the education field find quick hints regarding students' factors based on literature. Even though the researcher still needs to read each literature, this method might lighten the quick review task. The implication for this study is the students' actors are interrelated with each other. If one component of a student is not supported, it might affect another component and affect students' academic achievement. For future studies, all possible students' factors will be collected and built into a dataset, then a model capable of making recommendations will be developed.

6 Acknowledgement

This study is partly funded by Lembaga Pengelola Dana Pendidikan (LPDP). Ms Yunita is in her doctoral study supported by Beasiswa Unggulan Dosen Indonesia Dalam Negeri (BUDI-DN), Lembaga Pengelola Dana Pendidikan (LPDP), and cooperation of the Ministry of Research and Higher Education and the Ministry of Finance of the Republic of Indonesia.

7 References

[1] J. Anuradha, "A brief introduction on big data 5Vs characteristics and hadoop technology," in *Procedia Computer Science*, 2015, vol. 48, no. C, pp. 319–324. https://doi.org/10.1016/j.procs.2015.04.188

[2] S. Sedkaoui and M. Khelfaoui, "Understand, develop and enhance the learning process with big data," *Inf. Discov. Deliv.*, vol. 47, no. 1, pp. 2–16, 2019. https://doi.org/10.1108/IDD-09-2018-0043

[3] R. Al-Shabandar, A. J. Hussain, P. Liatsis, and R. Keight, "Analyzing learners behavior in MOOCs: An examination of performance and motivation using a data-driven approach," *IEEE Access*, vol. 6, pp. 73669–73685, 2018. https://doi.org/10.1109/ACCESS.2018.2876755

[4] D. A. Filvà, M. A. Forment, F. J. García-Peñalvo, D. F. Escudero, and M. J. Casañ, "Clickstream for learning analytics to assess students' behavior with Scratch," *Futur. Gener. Comput. Syst.*, vol. 93, pp. 673–686, 2019. https://doi.org/10.1016/j.future.2018.10.057

[5] D. J. Lemay and T. Doleck, "Grade prediction of weekly assignments in MOOCS: mining video-viewing behavior," *Educ. Inf. Technol.*, vol. 25, no. 2, pp. 1333–1342, 2020. https://doi.org/10.1007/s10639-019-10022-4

[6] M. Cantabella, R. Martínez-España, B. Ayuso, J. A. Yáñez, and A. Muñoz, "Analysis of student behavior in learning management systems through a Big Data framework," *Futur. Gener. Comput. Syst.*, vol. 90, pp. 262–272, 2019. https://doi.org/10.1016/j.future.2018.08.003

[7] M. Ocaña, H. Khosravi, and A. Bakharia, "Profiling language learners in the big data era," in *ASCILITE 2019-Conference Proceedings-36th International Conference of Innovation, Practice and Research in the Use of Educational Technologies in Tertiary Education: Personalised Learning. Diverse Goals. One Heart*, 2019, pp. 285–287.

[8] A. Y. Q. Huang, O. H. T. Lu, J. C. H. Huang, C. J. Yin, and S. J. H. Yang, "Predicting students' academic performance by using educational big data and learning analytics: evaluation of classification methods and learning logs," *Interact. Learn. Environ.*, vol. 28, no. 2, pp. 206–230, 2020. https://doi.org/10.1080/10494820.2019.1636086

[9] S. Qu, K. Li, S. Zhang, and Y. Wang, "Predicting achievement of students in smart campus," *IEEE Access*, vol. 6, pp. 60264–60273, 2018. https://doi.org/10.1109/ACCESS.2018.2875742

[10] V. Dhanalakshmi, D. Bino, and A. M. Saravanan, "Opinion mining from student feedback data using supervised learning algorithms," in *2016 3rd MEC International Conference on Big Data and Smart City, ICBDSC 2016*, 2016, pp. 332–336. https://doi.org/10.1109/ICBDSC.2016.7460390

[11] G. Elia, G. Solazzo, G. Lorenzo, and G. Passiante, "Assessing learners' satisfaction in collaborative online courses through a big data approach," *Comput. Human Behav.*, vol. 92, pp. 589–599, 2019. https://doi.org/10.1016/j.chb.2018.04.033

[12] R. K. Jena, "Sentiment mining in a collaborative learning environment: capitalising on big data," *Behav. Inf. Technol.*, vol. 38, no. 9, pp. 986–1001, 2019. https://doi.org/10.1080/0144929X.2019.1625440

[13] J. M. Lodge, S. S. J. Alhadad, M. J. Lewis, and D. Gašević, "Inferring learning from big data: The importance of a transdisciplinary and multidimensional approach," *Technol. Knowl. Learn.*, vol. 22, no. 3, pp. 385–400, 2017. https://doi.org/10.1007/s10758-017-9330-3

[14] D. Harris, "Factors affecting college grades: a review of the literature, 1930-1937," *Psychol. Bull.*, vol. 37, no. 3, pp. 125–166, 1940. https://doi.org/10.1037/h0055365

[15] X. Xiong, S. Zhao, E. G. Van Inwegen, and J. E. Beck, "Going deeper with deep knowledge tracing," in *Proceedings of the 9th International Conference on Educational Data Mining, EDM 2016*, 2016, pp. 545–550.

[16] T. Priatna, D. Maylawati, H. Sugilar, and M. Ramdhani, "Key success factors of e-learning implementation in higher education," *Int. J. Emerg. Technol. Learn.*, vol. 15, no. 17, pp. 101–114, 2020. https://doi.org/10.3991/ijet.v15i17.14293

[17] P. S. Muljana and T. Luo, "Factors contributing to student retention in online learning and recommended strategies for improvement: A systematic literature review," *J. Inf. Technol. Educ. Res.*, vol. 18, pp. 19–57, 2019. https://doi.org/10.28945/4182

[18] A. Rahmah, H. B. Santoso, and Z. A. Hasibuan, "Conceptualizing technology-enhanced learning constructs: A journey of seeking knowledge using literature-based discovery," in *Advances in Intelligent Systems and Computing*, vol. 1228 AISC, K. Arai, S. Kapoor, and R. Bhatia, Eds. Springer, Cham, 2020, pp. 746–759. https://doi.org/10.1007/978-3-030-52249-0_50

[19] G. Siemens, "Learning Analytics: The Emergence of a Discipline," *Am. Behav. Sci.*, vol. 57, no. 10, pp. 1380–1400, 2013. https://doi.org/10.1177/0002764213498851

[20] J. T. Avella, M. Kebritchi, S. G. Nunn, and T. Kanai, "Learning analytics methods, benefits, and challenges in higher education: A systematic literature review," *Online Learn.*, vol. 20, no. 2, pp. 13–29, 2016. https://doi.org/10.24059/olj.v20i2.790

[21] A. Yunita, H. B. Santoso, and Z. A. Hasibuan, "Research review on big data usage for learning analytics and educational data mining: A way forward to develop an intelligent automation system," *J. Phys. Conf. Ser.*, vol. 1898, no. 1, pp. 0–13, 2021. https://doi.org/10.1088/1742-6596/1898/1/012044

[22] M. Salihoun, "State of art of data mining and learning analytics tools in higher education," *Int. J. Emerg. Technol. Learn.*, vol. 15, no. 21, pp. 58–76, 2020. https://doi.org/10.3991/ijet.v15i21.16435

[23] M. Allahyari *et al.*, "A brief survey of text mining: Classification, clustering and extraction techniques," in *Proceedings of KDD Bigdas*, 2017.

[24] Y. B. Touimi, A. Hadioui, N. El Faddouli, and S. Bennani, "Intelligent chatbot-LDA recommender system," *Int. J. Emerg. Technol. Learn.*, vol. 15, no. 20, pp. 4–20, 2020. https://doi.org/10.3991/ijet.v15i20.15657

[25] B. Kitchenham, O. Pearl Brereton, D. Budgen, M. Turner, J. Bailey, and S. Linkman, "Systematic literature reviews in software engineering - A systematic literature review," *Inf. Softw. Technol.*, vol. 51, no. 1, pp. 7–15, 2009. https://doi.org/10.1016/j.infsof.2008.09.009

[26] P. Nuankaew, W. Nuankaew, D. Teeraputon, K. Phanniphong, and S. Bussaman, "Prediction model of student achievement in business computer disciplines," *Int. J. Emerg. Technol. Learn.*, vol. 15, no. 20, p. 160, 2020. https://doi.org/10.3991/ijet.v15i20.15273

[27] A. Yunita, H. B. Santoso, and Z. A. Hasibuan, "Deep learning for predicting students' academic performance," in *Proceedings of 2019 4th International Conference on Informatics and Computing, ICIC 2019*, 2019, pp. 1–6. https://doi.org/10.1109/ICIC47613.2019.8985721

[28] Y. Zhang, A. Ghandour, and V. Shestak, "Using learning analytics to predict students performance in moodle LMS," *Int. J. Emerg. Technol. Learn.*, vol. 15, no. 20, pp. 102–114, 2020. https://doi.org/10.3991/ijet.v15i20.15915

[29] A. Martín-Martín, E. Orduna-Malea, M. Thelwall, and E. Delgado López-Cózar, "Google Scholar, Web of Science, and Scopus: A systematic comparison of citations in 252 subject categories," *J. Informetr.*, vol. 12, no. 4, pp. 1160–1177, 2018. https://doi.org/10.1016/j.joi.2018.09.002

[30] N. J. van Eck and L. Waltman, "Software survey: VOSviewer, a computer program for bibliometric mapping," *Scientometrics*, vol. 84, no. 2, pp. 523–538, 2010. https://doi.org/10.1007/s11192-009-0146-3

[31] N. J. van Eck and L. Waltman, "Text mining and visualization using VosViewer," *Text Min. Vis.*, 2011.

[32] A. Kumar and P. K. Saini, "Need of Facet Analysis of MOOCs : A Review and Feasibility Study," *Libr. Philos. Pract.*, pp. 9–15, 2021.

[33] V. Broughton, "The need for a faceted classification as the basis of all methods of information retrieval," in *Aslib proceedings*, 2006. https://doi.org/10.1108/00012530610648671

[34] B. Hjørland, "Facet analysis: The logical approach to knowledge organization," *Inf. Process. \& Manag.*, vol. 49, no. 2, pp. 545–557, 2013. https://doi.org/10.1016/j.ipm.2012.10.001

[35] D. Kim, M. Yoon, I. H. Jo, and R. M. Branch, "Learning analytics to support self-regulated learning in asynchronous online courses: A case study at a women's university in South Korea," *Comput. Educ.*, vol. 127, no. August, pp. 233–251, 2018. https://doi.org/10.1016/j.compedu.2018.08.023

[36] T. Nethanomsak, T. K. Ngang, and S. Raksasataya, "Gender issues in education: Why boys do poorly," *Pertanika J. Soc. Sci. Humanit.*, vol. 27, no. T2, pp. 119–138, 2019.

[37] Y. Zhang, X. Qin, and P. Ren, "Adolescents' academic engagement mediates the association between Internet addiction and academic achievement: The moderating effect of classroom achievement norm," *Comput. Human Behav.*, vol. 89, no. January, pp. 299–307, 2018. https://doi.org/10.1016/j.chb.2018.08.018
[38] T. Fernández-Villa, A. J. Molina, M. García-Martín, J. Llorca, M. Delgado-Rodríguez, and V. Martín, "Validation and psychometric analysis of the Internet Addiction Test in Spanish among college students," *BMC Public Health*, vol. 15, no. 1, 2015. https://doi.org/10.1186/s12889-015-2281-5
[39] X. S. Wu *et al.*, "Prevalence of Internet addiction and its association with social support and other related factors among adolescents in China," *J. Adolesc.*, vol. 52, pp. 103–111, 2016. https://doi.org/10.1016/j.adolescence.2016.07.012
[40] I. Han and W. S. Shin, "The use of a mobile learning management system and academic achievement of online students," *Comput. Educ.*, vol. 102, pp. 79–89, 2016. https://doi.org/10.1016/j.compedu.2016.07.003
[41] Z. guo Yin, "Improving learners' cognitive adjustment in a network multimedia learning environment," *World Trans. Eng. Technol. Educ.*, vol. 14, no. 1, pp. 83–88, 2016.
[42] G. K. Y. Lee, "Student and teacher perspectives on factors to success and attrition in engineering mini-mooc using technology acceptance model (TAM)," *Humanit. Soc. Sci. Rev.*, vol. 7, no. 1, pp. 310–315, 2019. https://doi.org/10.18510/hssr.2019.7135
[43] P. Ifinedo, J. Pyke, and A. Anwar, "Business undergraduates' perceived use outcomes of Moodle in a blended learning environment: The roles of usability factors and external support," *Telemat. Informatics*, vol. 35, no. 1, pp. 93–102, 2018. https://doi.org/10.1016/j.tele.2017.10.001
[44] Ž. Namestovski, L. Major, G. Molnár, Z. Szűts, P. Esztelecki, and G. Kőrösi, "External motivation, the key to success in the MOOCs framework," *Acta Polytech. Hungarica*, vol. 15, no. 6, pp. 125–142, 2018.
[45] S. D. Teasley, "Student facing dashboards: One size fits all?," *Technol. Knowl. Learn.*, vol. 22, no. 3, pp. 377–384, 2017. https://doi.org/10.1007/s10758-017-9314-3
[46] E. Hovdhaugen, "Working while studying: the impact of term-time employment on dropout rates," *J. Educ. Work*, vol. 28, no. 6, pp. 631–651, 2015.
[47] Z. Siddique, P. L. Hardré, and D. Altan, "Effects of a mechanical engineering design course on students' motivational features," *Int. J. Mech. Eng. Educ.*, vol. 43, no. 1, pp. 44–74, 2015. https://doi.org/10.1177/0306419015581734
[48] M. Wu, I. Siswanto, and C. Ko, "The influential factors and hierarchical structure of college students' creative capabilities—An empirical study in Taiwan," *Think. Ski. Creat.*, vol. 26, pp. 176–185, 2017. https://doi.org/10.1016/j.tsc.2017.10.006
[49] D. Y. Jacobsen, "Dropping out or dropping in? A connectivist approach to understanding participants' strategies in an e-learning MOOC pilot," *Technol. Knowl. Learn.*, vol. 24, no. 1, pp. 1–21, 2019. https://doi.org/10.1007/s10758-017-9298-z
[50] C. Costa-Santos, P. Vieira-Marques, A. Costa-Pereira, M. A. Ferreira, and A. Freitas, "Do students from public schools fare better in medical school than their colleagues from private schools? if so, what can we learn from this?," *BMC Med. Educ.*, vol. 18, no. 1, 2018. https://doi.org/10.1186/s12909-018-1162-x
[51] A. Khan and S. K. Ghosh, "Data mining based analysis to explore the effect of teaching on student performance," *Educ. Inf. Technol.*, vol. 23, no. 4, pp. 1677–1697, 2018. https://doi.org/10.1007/s10639-017-9685-z
[52] L. M. Jeno, J. A. Grytnes, and V. Vandvik, "The effect of a mobile-application tool on biology students' motivation and achievement in species identification: A Self-Determination Theory perspective," *Comput. Educ.*, vol. 107, pp. 1–12, 2017. https://doi.org/10.1016/j.compedu.2016.12.011

[53] M. A. Nkrumah, "The relevance of teacher factors in understanding tertiary students' performances," *Qual. Assur. Educ.*, vol. 26, no. 4, pp. 476–488, 2018. https://doi.org/10.1108/QAE-02-2018-0017
[54] S. Sivarajah, S. M. Smith, and S. C. Thomas, "Tree cover and species composition effects on academic performance of primary school students," *PLoS One*, vol. 13, no. 2, 2018. https://doi.org/10.1371/journal.pone.0193254
[55] D. A. Cook, B. L. Gas, and A. R. Artino, "Measuring mindsets and achievement goal motivation: A validation study of three instruments," *Acad. Med.*, vol. 93, no. 9, pp. 1391–1399, 2018. https://doi.org/10.1097/ACM.0000000000002290
[56] M. S. Kadir, A. S. Yeung, and T. M. O. Diallo, "Simultaneous testing of four decades of academic self-concept models," *Contemp. Educ. Psychol.*, vol. 51, pp. 429–446, 2017. https://doi.org/10.1016/j.cedpsych.2017.09.008
[57] L. Yáñez-Marquina and L. Villardón-Gallego, "Attitudes towards mathematics at secondary level: Development and structural validation of the Scale for Assessing Attitudes towards Mathematics in Secondary Education (SATMAS)," *Electron. J. Res. Educ. Psychol.*, vol. 14, no. 3, pp. 557–581, 2016. https://doi.org/10.14204/ejrep.40.15163
[58] E. Panadero, "A review of self-regulated learning: Six models and four directions for research," *Front. Psychol.*, vol. 8, no. APR, pp. 1–28, 2017. https://doi.org/10.3389/fpsyg.2017.00422
[59] M. Te Wang, J. Guo, and J. L. Degol, "The role of sociocultural factors in student achievement motivation: A cross-cultural review," *Adolesc. Res. Rev.*, vol. 5, no. 4, pp. 435–450, 2020. https://doi.org/10.1007/s40894-019-00124-y
[60] O. Shapira-Lishchinsky and M. Ben-Amram, "Exploring the social ecological model based on national student achievements: extracting educational leaders' role," *Int. J. Leadersh. Educ.*, vol. 21, no. 3, pp. 380–398, 2018. https://doi.org/10.1080/13603124.2017.1318956
[61] A. Fortin, L. Sauvé, C. Viger, and F. Landry, "Nontraditional student withdrawal from undergraduate accounting programmes: a holistic perspective," *Account. Educ.*, vol. 25, no. 5, pp. 437–478, 2016. https://doi.org/10.1080/09639284.2016.1193034
[62] M. Shannon, C. M. N. Barry, A. DeGrace, and T. DiDonato, "How parents still help emerging adults get their homework done: The role of self-regulation as a mediator in the relation between parent--child relationship quality and school engagement," *J. Adult Dev.*, vol. 23, no. 1, pp. 36–44, 2016. https://doi.org/10.1007/s10804-015-9219-0
[63] A. B. Duche-Perez, F. M. Paredes-Quispe, and O. A. Gutierrez-Aguilar, "The Transition from high school to university: Identifying internal and external factors for a successful transition in peruvian students of Architecture and Engineering," in *EDUNINE 2019 - 3rd IEEE World Engineering Education Conference: Modern Educational Paradigms for Computer and Engineering Career, Proceedings*, 2019. https://doi.org/10.1109/EDUNINE.2019.8875751
[64] F. T. C. Schmidt, J. Fleckenstein, J. Retelsdorf, L. Eskreis-Winkler, and J. Möller, "Measuring grit: A German validation and a domain-specific approach to grit," *Eur. J. Psychol. Assess.*, vol. 35, no. 3, pp. 436–447, 2019. https://doi.org/10.1027/1015-5759/a000407
[65] M. H. Shin, "How to use artificial intelligence in the english language learning classroom," *Indian J. Public Heal. Res. Dev.*, vol. 9, no. 9, pp. 557–561, 2018. https://doi.org/10.5958/0976-5506.2018.01058.6
[66] P. E. Brockmann, M. Schlaud, C. F. Poets, and M. S. Urschitz, "Predicting poor school performance in children suspected for sleep-disordered breathing," *Sleep Med.*, vol. 16, no. 9, pp. 1077–1083, 2015. https://doi.org/10.1016/j.sleep.2015.03.021

[67] M. B. Kerby, "Toward a new predictive model of student retention in higher education: An application of classical sociological theory," *J. Coll. Student Retent. Res. Theory Pract.*, vol. 17, no. 2, pp. 138–161, 2015. https://doi.org/10.1177/1521025115578229
[68] S. Scherer, C. P. Talley, and J. E. Fife, "How personal factors influence academic behavior and GPA in African American STEM students," *SAGE Open*, vol. 7, no. 2, 2017. https://doi.org/10.1177/2158244017704686
[69] Y. Alas, M. Anshari, N. I. Sabtu, and N. Yunus, "Second-chance university admission, the theory of planned behaviour and student achievement," *Int. Rev. Educ.*, vol. 62, no. 3, pp. 299–316, 2016. https://doi.org/10.1007/s11159-016-9558-5
[70] S. Burgis-Kasthala, N. Elmitt, L. Smyth, and M. Moore, "Predicting future performance in medical students. A longitudinal study examining the effects of resilience on low and higher performing students," *Med. Teach.*, vol. 41, no. 10, pp. 1184–1191, 2019. https://doi.org/10.1080/0142159X.2019.1626978

8 Authors

Ariana Yunita is a doctoral student in the Faculty of Computer Science, Universitas Indonesia, and a lecturer at Department of Computer Science, Universitas Pertamina, Indonesia. She received bachelor degree in Informatics Engineering from Institut Teknologi Sepuluh Nopember, Indonesia, 2009 and master degree in Information Technology from James Cook University, Australia, 2015. Her research interests include Artificial Intelligence in Education, Big Data, Data Science, Text Mining, Machine Learning, Learning Analytics and Educational Data Mining (e-mail: ariana.yunita@universitaspertamina.ac.id).

Harry Budi Santoso, PhD is an associate professor at the Faculty of Computer Science, Universitas Indonesia (UI). He received his B.Sc. and M.Sc. in Computer Science from UI, and his PhD in Engineering Education from the Department of Engineering Education at Utah State University, USA. At the national level, he participates in a professional organization "Association for Higher Education Informatics and Computers (APTIKOM)" as the Head of the Lecturer Competence Division. He is also currently the Head of Digital Library and Distance Learning lab at the Faculty of Computer Science Universitas Indonesia. His research has been published in scientific journals and international conference proceedings indexed by Scopus, IEEE Xplore, ACM Digital Library, Springer, and Web of Science. He has written more than 150 Scopus indexed scientific papers. His research interest includes learning personalization, engineering and computer science education, metacognition and self-regulated learning, human-computer interaction, user experience, and online distance learning.

Zainal A. Hasibuan, PhD is a full professor in computer science and information system at Universitas Dian Nuswantoro. He received BSc. degree in Statistic and Computation from Bogor Institute of Agriculture (IPB University), Indonesia, 1986, MSc. and PhD in Information Science, Indiana University, in 1989 and 1995 respectively. Currently, he is a lecturer at Faculty of Computer Science, Universitas Dian Nuswantoro. He is also the chairman of Association of Higher Education in Informatics and Computer (APTIKOM), Indonesia and member of National ICT Council. His research interests include Artificial Intelligence, Big Data, Data Science, e-Learning,

Digital Library, Information Retrieval, Information System, and Software Engineering. He has written more than 200 scientific papers in the form of books, papers in various seminars, international conferences, local and international journals.

Article submitted 2022-04-25. Resubmitted 2022-06-11. Final acceptance 2022-06-13. Final version published as submitted by the authors.

9 Appendix: Terms of clusters

Table 1. List of terms in each cluster

Cluster	List of Terms				
Cluster 1 – 39 terms	academic performance; adaptation; adolescent; age; association; behaviour; boy; challenge;	Child; college student; comparison; content; country; demography characteristics; difference; domain;	Efficiency; Evaluation; Family; Feeling; Further study; Gender; Girl; Goal;	Gpa; high school; student; internal consistency; measure; parent; participant; predictor;	psychometric property; relation; reliability; self-regulation; strength; structure; student engagement
Cluster 2- 35 terms	Access; Analytic; Attrition; Business; College; Control; Course; current study; engagement; external factor;	extrinsic motivation; first year; information; interaction; internal consistency reliability; internal factor; lack;	learning process; motivational factor; outcome; population; retention; science; self-efficacy; series;	student retention; students performance; technology; time; usefulness; work	Access; Analytic; Attrition; Business; College; Control; Course; current study; engagement; external factor;
Cluster 3 – 32 terms	Activity; Administrator; Assessment; Class; Classroom; Competence; Educator; Effectiveness;	Example; Expectation; external motivation; feedback; higher education; importance;	Improvement; Individual; Instructor; intrinsic motivation; medium; need; perspective;	Question; Recommendation; Situation; student achievement; student; motivation;	Support; Task; Teacher; Teaching; Variance; Weakness;
Cluster 4 – 30 terms	Aspect; Attitude; Background; Culture; Difficulty; Discussion;	doctoral student; effort; examination; experience; faculty; intervention;	Issue; Literature; male student; nature; opportunity; perception;	Persistence; Practice; Process; Respect; Role; Satisfaction;	skill strategy success term transition understanding
Cluster 5 – 14 words	Ability; academic achievement; Area;	Behaviour; Body; Computer; Intelligence;	Interest; Mathematics; Personality;	statistical analysis; students academic achievement;	subject; value;