Journal of Computers Vol. 34 No. 4, August 2023, pp. 233-240
doi: 10.53106/199115992023083404020

An Analysis of Online Learner Types Applicable to Lifelong Learning Environments

Shiow-Lin Hwu[1*], Sheng-Lung Peng[2]

[1] Department of Digital Multimedia Design, National Taipei University of Business, Taoyuan 324, Taiwan, ROC
slinghu@ntub.edu.tw

[2] Department of Creative Technologies and Product Design, National Taipei University of Business, Taoyuan 324, Taiwan, ROC
slpeng@ntub.edu.tw

Received 1 April 2023; Revised 1 May 2023; Accepted 14 June 2023

Abstract. Based on the perspective of Sustainable Development Goals (SDGs) Quality Education and lifelong learning, it is necessary to respect the learning opportunities and quality for all individuals. Online learning can provide more opportunities for lifelong learning, but due to the significant differences in students' backgrounds and characteristics, personalized and timely support becomes more crucial. Learning analytics (LA) in online learning environment is a way to facilitate understanding of the potential meaningful information and relationships of students. One of the main functions of LA is to monitor the learning performance and identify potential learning problems early. In this study, k-means clustering is performed to determine the types of learning in lifelong online learning environments, based on students' personal traits (background factors), learning behavior paths, and interactive perspectives on learning performance. Moreover, statistical analysis is used to further evaluate the linear correlation coefficients as well as the characteristics of each group of students, who ranged in age from 18 to 73, with a total of 2386 participants from five courses, in the interactive perspective. The result shows a significant correlation between learning performance and persistence across the three learning clusters, with a tendency towards continuous learning, thus providing educators an understanding of the learning behavior characteristics of those types of online learners.

Keywords: k-means clustering, learning analytics, lifelong online learning, SDGs

1 Introduction

The United Nations' Sustainable Development Goals (SDGs) emphasize the importance of quality education and lifelong learning as Goal 4 out of 17 interconnected and integrated goals [1]. This goal aims to ensure that all individuals have access to inclusive and equitable quality education and lifelong learning opportunities. However, promoting lifelong learning is not without challenges, as it requires addressing diverse student characteristics, flexible learning environments, and difficulties in monitoring teaching effectiveness. Online learning presents a well solution, as it can effectively overcome the limitations of physical space and time. It also provides access to a wealth of learning resources on the internet, catering to the unique needs and characteristics of both educators and learners.

Due to the outbreak of the COVID-19 pandemic, education over the world has rapidly transitioned to online learning on a global scale. The use of e-learning technology has increased significantly since it can easily offer greater interactivity and productivity in the teaching and learning process. Online learning is a solution to guarantee effective instruction and teachers have to monitor the leaerning effectiveness of all learners, particularly young students who are more susceptible to distractions.

The effectiveness of online learning, however, has been a challenge for educators and researchers. In recent years, one of the increasingly popular approach is learning analytics in online learning environments [2-4]. The approach involves collection of large datasets (i.e. big data) regarding of students' learning processes, either through widely-used traditional learning management platforms such as Moodle and Blackboard, or through custom-designed virtual devices [2]. As claimed, the goal of online learning focuses on the improvement of online learning quality.

* Corresponding Author

The use of complex algorithms to analyze students' digital footprints offers a promising avenue for improving online learning experiences [5]. Through the examination of these digital footprints, insights can be gained into how students participate in various online learning activities [3]. This, in turn, can achieve a deeper understanding of the factors that contribute to successful learning outcomes [3, 6-8]. The primary objective of this approach [9] is to identify students' learning status at an early stage and provide timely and personalized support. By detecting patterns in students' digital footprints, educators can offer interventions that address potential issues before they become significant obstacles to academic success.

However, while the analysis of digital footprints provides valuable information, some scholars advocate for a more comprehensive approach to student data analysis. In addition to behavioral data, academic performance, learning and teaching behavior, and socioeconomic status should also be considered to provide a more holistic understanding of students' learning experiences [3, 9]. This multifaceted approach can facilitate the development of personalized and supportive interventions that meet the unique needs of individual learners.

In conclusion, the analysis of students' digital footprints has the potential to revolutionize online learning by providing valuable insights into students' participation in various learning activities. By considering a range of student data, educators and administrators can develop personalized and supportive interventions that foster academic success and improve the overall learning experience.

Guo et al. [10] conducted a thorough analysis of the learning behaviors of online learners and developed visual representations called learner portraits to depict the hidden behavioral characteristics of learners during the online learning process. Their study was based on the tracking and analysis of online learning behavior data, which was used to construct learner portraits. While this approach was effective in characterizing the online learning behaviors of students, it did not comprehensively consider the overall data of students and the diverse learning environments of lifelong learners. As a result, the personalized supportive interventions that can be provided by learner portraits are relatively limited. This limitation may hinder the ability of educators and institutions to provide effective support to students with diverse learning needs.

The objective of this study is to investigate the learning behavior characteristics of different types of online learners from an interactive perspective that considers students' personal characteristics, learning behavior paths, and learning outcomes.

The following research questions will guide the investigation:

1.What are the distinct types of students' learning styles?

2.What are the underlying behavioral characteristics of different types of students during the online learning process?

2 Literature Review

In this research, one main purpose is to provide educators with a comprehensive understanding of the unique learning behavior characteristics of different types of online learners. As such, this section describes a brief overview of relevant research in the field of learning analytics.

The integration of online learning modalities has become a pervasive trend in educational contexts. Online learning platforms digitize the process of teacher-student interaction and instruction. Although learning data has been used in educational statistics, the growth in the quantity and types of data has led to a greater focus on the importance of learning analytics [2, 11, 12]. Compared to traditional questionnaire-based statistical analysis methods, analyzing online learning behavior can provide various data related to online learning, resulting in more objective data than subjective responses typically obtained through questionnaires. The emergence of learning analytics has brought benefits to the educational field, as it can be used to analysis authentic data from students to identify the problems encountered in e-learning and to provide intervention to assist students [3, 6].

But the use of data collected from technology mediated interactions and its use in algorithms is often insufficient to gain a comprehensive vision of a learning experience [2]. A comprehensive understanding of personalized learning quality can be achieved through the integration of education, learning science, and computer science research fields, taking a multidimensional approach [2, 12].

Clustering methods are used to discover the relationship between students' behaviors and their learning performance [14, 15]. The application of various clustering algorithm has been applied in many a case to educational data set in diverse studies. Such as, the interactions of on-line learners are clustered by the hierarchical clustering algorithm and k-means algorithm in order to discover the relationship between the final grades and the use of the modules [16]. The k-means clustering method has been applied to analyze the changes in self-assessed skills before and after their submission, which provides insight into the role of self-assessment in determining final grades [17]. In another approach, Francis and Babu [18] classified students into three groups (high, medium, low) by using four popular classifiers (SVM, Naive Bayes, Decision Tree, and Neural network). After that, k-mean clustering and majority voting are used to predict the best accuracy of students. It is apparent that k-means clustering has become a prevalent technique employed in the field of learning analytics.

Trajectory behaviors and learning behaviors is different. The difference between the trajectory behavior and the resource learning behavior is that the resource learning behavior contains the body of the search, and the track behavior is only to retrieve the action, not including the main content of the retrievation [13]. In learning analysis, the discussion should include a series of actions during the learning process, rather than just learning trajectories, by treating each learning behavior as a system.

In summary, comprehensive exploration and analysis are necessary. Not only should we investigate students' personal background characteristics as a factor influencing learning achievement, but also learning behaviors and academic performance should be included in the scope of learning analysis.

3 Proposed Method

In this section, an explanation of data collection and analysis dimensions will be presented. In addition, we will describe an analytical framework and methods for each group to illustrate the characteristics of different types of learners.

3.1 Data Collection and Dimensions

This study analyzed from the university-level students who were exclusively enrolled in online courses. The main objective of the university focuses on promote of lifelong learning, resulting in a diverse student population in terms of age and occupation. The curriculum consisted of five courses, two of which were in the field of science and technology, while the other three were related to management. The students ranged in age from 18 to 73, with a total of 2386 participants. The learning data is collected over the course of four semesters since 2019.

The data sources for this study include three major types: learner characteristics, learning behavior pathways, and learning performance. Learner characteristics include two parts: personality and environment. This indicates that the learner's characteristics are shaped by both their individual traits and their surrounding circumstances.

The learning behavior pathways were divided into two categories, system interaction behaviors and resource interaction behaviors, with reference to the literature [10]. And the Learning Performance consists of two main components: formative assessment and summative assessment. Formative assessment is used during the learning process to provide feedback to students and help them improve their understanding and skills. It is usually ongoing and focuses on individual progress. Summative assessment, on the other hand, is used to evaluate student learning at the end of a unit, course, or program. It is usually a final assessment and measures the overall level of understanding and achievement. Learning persistence represents a student's ability to complete assignments and exams during the semester, and the data was divided into completed and non-completed categories. Table 1 summarizes the data and its characteristic values.

In the study [10], the system interaction behaviors refer to the interactions between learners and the learning management system, while the resource interaction behaviors refer to the interactions between learners and learning resources, such as online materials and peer interactions.

Table 1. Classification of online learning behaviour

Data source types	Subtypes	Characteristic data
Learner characteristics	Personality	Age, occupation
	Environment	Major, educational background upon enrollment
Learning behavior pathways	System interaction behaviors	Number of class attendance
	resource interaction behaviors	Learning progress, reading hours
Learning performance	Formative assessment	Regular course grades, final exam grades
	Summative assessment	Semester grades, learning persistence

3.2 Analysis Methods

The purpose of this study is to analyze the multi-dimensional data of online learning students in order to understand the student types in online learning, and to explore the interactive relationships among personal background characteristics, learning behavior paths, and learning performance for each student type. To achieve this goal, we will first use clustering methods to conduct dimensionality reduction analysis on the online learning data. Clustering is a method of categorizing data into groups, which is an unsupervised learning approach, meaning that there are no pre-defined labels for the training data. Its main purpose is to identify several clusters of similar data, so that members in the same subset have similar attributes, commonly by using the k-means clustering algorithm to find shorter spatial distances in the coordinate system.

After clustering the reduced-dimension data, different learning types are identified. In order to depict the characteristics of different types of learners, statistical methods will then be used to investigate the correlation coefficients and variance analysis of various dimension indicators of individual background characteristics, learning behavior paths, and learning performance in each different learning type. The research framework for exploring the relationship between various data is as Fig. 1:

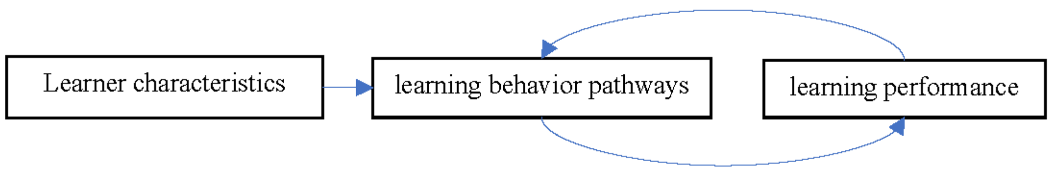

Fig. 1. Data analytical framework after clustering

4 Results and Discussions

This section presents the analysis results and provides a detailed classification and description of the groups. We classified online learners into three major groups based on our analysis, and the classification method and characteristics of each group are explained in this chapter.

4.1 Result of Classification

Firstly, this study utilized the k-means clustering algorithm to understand the learning types in the online learning environment. Based on the result of operating the elbow method (Fig. 2), a significant inflection point was found at k=3, indicating that the optimal value of k is 3 and the clustering effect is best when the number of clusters is 3. Three clusters are formed, forming three learning groups. Secondly, this study differentiated different types of learners by conducting mean analysis on the characteristic values of the three groups of learners (Table 2).

Derived from Table 2, we found that three groups of students are relatively homogeneous in terms of personal background characteristics and learning performance, but there are significant differences in learning behavior paths. Therefore, based on the average values of learning behavior paths, the three groups of students were named as follows: Group 2 as Highly Engaged, Group 3 as Moderately Engaged, and Group 1 as Lowly Engaged.

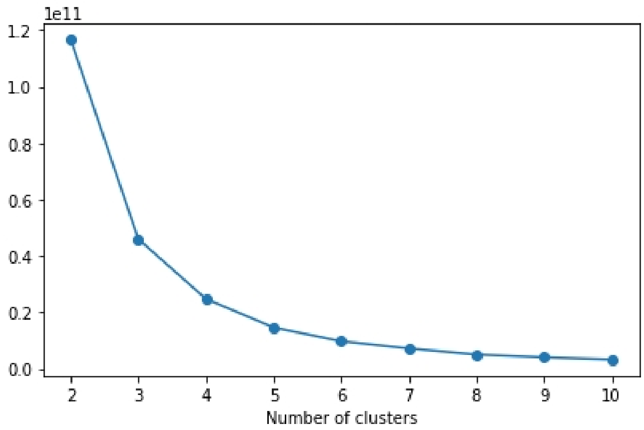

Fig. 2. The elbow method

Table 2. Mean values of group values

Group	Named	Students	Learner characteristics	Learning behavior pathways	Learning performance
1	Lowly engaged	2052	14.32	261.99	56.93
2	Highly engaged	120	17.09	21388.65	62.33
3	Moderately engaged	214	16.05	8611.06	59.27

4.2 Result of Highly Engaged Group (Group2)

The Highly Engaged group of students had the smallest number of students among the three groups (Table 2). Their learning behavior pathways had the highest values for various indicators, such as class attendance, learning progress, and reading hours, compared to the other two groups. Therefore, this study named this group the Highly Engaged group of learners.

Multivariate analysis of variance showed that personal traits and environmental traits in the learner characteristics significantly affected learning behavior. In terms of learning progress, different age learners had different levels of interaction with online materials in the Highly Engaged group. Mature students had higher reading progress. In addition, there was a significant positive correlation between educational background and reading hours, indicating that the higher the educational background, the higher the reading hours.

The information in the lower triangular matrix of Table 3 to Table 5 is redundant and can be omitted as it is identical to the information in the upper triangular matrix. The F value in the table corresponds to the test statistic of the multivariate analysis of variance (MANOVA), which indicates the presence of a significant difference between the analyzed variables across different groups. The r value in the table represents the correlation coefficient, which indicates the strength of the relationship between two variables.

Upon examining the data of this student group, moderate to low correlations were found between multiple feature data. In terms of the relationship between learning behavior and learning performance, except for the lack of significant correlation between class attendance and reading hours, and between learning progress and usual grades, all other correlations were significant. Additionally, learning progress, reading time, usual grades, final exam scores, and semester grades were found to have a significant positive correlation with whether students dropped out during the semester.

Table 3. Statistical summary table for the G2 Highly Engaged Group

Analysis item	Characteristic fata	Attendance	Learning progress	Reading hours	Regular course grades	Final exam grades	Semester grades
Learner characteristics and learning behavior pathways	Age	F=1.36	F=2.924**	F=1.775			
	Occupation	F=1.331	F=1.615	F=1.667			
	Major	F=1.40	F=0.89	F=93			
	Enrollment	F= 1.06	F= 1.21	F= 1.89*			
Learning behavior pathways and learning performance	Attendance	r=1	r=0.23*	r=0.35	r=0.22*	r=0.25**	r=0.27**
	Learning progress		r=1	r=0.27**	r=0.17	0.18*	0.24**
	Reading hours			r=1	r=0.21*	0.24**	0.28**
	Regular course grades				r=1	0.53**	0.79**
	Final exam grades					r=1	0.89**
	Semester grades						r=1
Learning persistence and learning performance	Learning persistence	r=0.10	r=0.24**	r=0.18*	r=0.60**	r=0.73**	r=0.77**

*<=.05 **<=.01

4.3 Result of Moderately Engaged Group (Group3)

The moderately engaged group (Table 2) represents the group with median values for both the number of students and the learning behavior pathway.

Regarding the impact of learner characteristics on learning behavior pathway, a multivariate analysis of variance showed that personal traits and environmental traits did not reach a significant level, indicating that it is difficult to distinguish differences in learning behavior among individuals with different background characteristics.

The significance of correlation coefficients was classified into three levels: high correlation (>=0.8), moderate correlation (0.4~0.8), and low correlation (<0.4). The correlation analysis in this group indicated that the number of class attendance was significantly and highly correlated with both learning progress and reading hours. Learning progress was moderately correlated with reading hours. It is noteworthy that all three grades of this group were significantly and highly correlated, indicating that the learning of this group is particularly manifested in their academic performance.

Table 4. Statistical summary table for the G3 Moderately Engaged Group

Analysis item	Characteristic fata	Attendance	Learning progress	Reading hours	Regular course grades	Final exam grades	Semester grades
Learner characteristics and learning behavior pathways	Age	F=0.64	F=1.52	F=0.63			
	Occupation	F=1.05	F=0.63	F=0.97			
	Major	F=1.06	F=0.97	F=1.15			
	Enrollment	F=0.56	F=1.57	F=1.31			
Learning behavior pathways and learning performance	Attendance	r=1	r=0.25 **	r=1.36*	r=0.64	r=0.11	r=0.12
	Learning progress		r=1	r=0.22**	r=0.78	r=0.29	r=0.61
	Reading hours			r=1	r=0.12	r=0.84	r=0.58
	Regular course grades				r=1	r=0.75**	r=0.80**
	Final exam grades					r=1	r=0.93**
	Semester grades						r=1
Learning persistence and learning performance	Learning persistence	r=0.61	r=0.62	r=0.48	r=0.71**	r=0.71**	r=0.83**

*<=.05 **<=.01

4.4 Result of Lowly Engaged Group (Group1)

The group of students with low learning engagement (named as such in this study) had the largest number of participants among all groups (Table 2). The values for class attendance, learning progress, and reading hours in their learning behavior pathway were the lowest among the three groups. Multivariate analysis of variance showed that both personal and environmental characteristics significantly affected learning behavior pathway, indicating that individuals with different learner characteristics exhibited significant differences in their learning behaviors pathway.

Regarding correlation analysis, class attendance was significantly related to learning progress and reading hours. Only in this lowly engaged group, learning progress was significantly related to regular grades. Similar to the moderate engaged group, all three grades of students in this group had significant and high correlations with each other, indicating that the learning behavior pathway of this group had a particular representation in terms of grades, and tended to exhibit a characteristic of sustained learning (learning persistence).

Table 5. Statistical summary table for the G1 Lowly Engaged Group

Analysis item	Characteristic fata	Attendance	Learning progress	Reading hours	Regular course grades	Final exam grades	Semester grades
Learner characteristics and learning behavior pathways	Age	F=4.85**	F=11.13**	F=12.52**			
	Occupation	F=4.885**	F=8.906**	F=9.461**			
	Major	F=5.53**	F=6.09**	F=11.39**			
	Enrollment	F= 2.38**	F= 7.89**	F= 9.19**			
Learning behavior pathways and learning performance	Attendance	r=1	r=0.28**	r=0.26**	r=0.24	r=-0.41	r=-0.01
	Learning progress		r=1	0.49	0.06**	-0.01	0.01
	Reading hours			r=1	0.04	-0.05	-0.02
	Regular course grades				r=1	0.81**	0.89**
	Final exam grades					r=1	0.95**
	Semester grades						r=1
Learning persistence and learning performance	Learning persistence	r=0.02	r=0.02	r=0.00	r=0.839**	r=0.830**	r=0.913**

*<.05 **<.01

5 Conclusion

The purpose of this investigation is to provide educators with a comprehensive understanding of the unique learning behavior characteristics of different types of online learners. By identifying the diverse learning styles and behaviors of students, educators can tailor their teaching strategies to better support the learning needs of their students. This research aims to contribute to the development of effective and personalized online learning environments, which are crucial in promoting student success in online learning contexts.

From the perspective of Sustainable Development Goals (SDGs) Quality Education and lifelong learning, it is imperative to ensure equal and quality learning opportunities for all individuals, regardless of their backgrounds. By examining the characteristics of different learning groups among lifelong learners from diverse backgrounds, teachers can gain valuable insights to better support their students.

In particular, the study identifies three distinct learning groups in the online learning environment of SDGs lifelong learning, each displaying unique patterns in terms of their learning characteristics, learning behaviors, and learning performances. Among these groups, the low-participation group stands out as the largest group in the lifelong learning environment, characterized by a relatively low level of engagement yet a strong inclination towards continuous learning. Notably, the study also reveals significant differences in learning participation among students with different background characteristics, highlighting the importance of understanding and addressing these differences to promote equitable learning opportunities for all.

The analysis reveals a significant correlation between learning performance and persistence across the three learning clusters, with a tendency towards continuous learning. This is encouraging as it suggests that learners possess a basic learning motivation and there is some supportive evidence for the sustainability of learning.

However, the results also indicate that learning performance and persistence levels vary across different levels of participation. Future research could examine formative assessment data to provide predictive and warning signals for learning sustainability.

The analysis also highlights the impact of learning background on the learning performance of low-participation students. In contrast, moderate and high-participation learners exhibit more active learning involvement, without any significant differences in their background characteristics. In viewpoint of these findings, future research should explore ways to decrease the population of learners with lowly engaged learner or provide tailored assistance to increase their engagement with the learning process.

Acknowledgement

The research presented in this paper was supported by a grant from the National Science Council, Taiwan (Grant No. MOST-110-2637-H-141-003). We would like to express our sincere appreciation to the council for their financial support, which made this study possible.

References

[1] United Nations, SDG 4: Ensure Inclusive and Equitable Quality Education and Promote Lifelong Learning Opportunities for All, United Nations, <https://sdgs.un.org/goals/goal4>, (accessed 01.16.23).
[2] A. Pardo, F. Han, R.A. Ellis, Combining university student self-regulated learning indicators and engagement with online learning events to predict academic performance, IEEE Transactions on Learning Technologies 10(1)(2017) 82-92.
[3] A.L. Sønderlund, E. Hughes, J. Smith, The efficacy of learning analytics interventions in higher education: A systematic review, British Journal of Educational Technology 50(5)(2019) 2594-2618.
[4] S.N. Kew, Z. Tasir, Developing a Learning Analytics Intervention in E-learning to Enhance Students' Learning Performance: A Case Study, Education and Information Technologies 27(5)(2022) 7099-7134.
[5] R.S. Baker, G. Siemens, Educational data mining and learning analytics, in: The Cambridge Handbook of the Learning Sciences, R. K. Sawyer (Ed.), Cambridge, U.K.: Cambridge University Press, 2014 (253-274).
[6] S.N. Kew, Z. Tasir, Developing a Learning Analytics Intervention in E-learning to Enhance Students' Learning Performance: A Case Study, Education and Information Technologies 27(5)(2022) 7099-7134.
[7] C. Perrotta, B. Williamson, The social life of Learning Analytics: cluster analysis and the 'performance' of algorithmic education, Learning, Media and Technology 43(1)(2018) 3-16.
[8] B. Williamson, Digital Methodologies of Education Governance: Pearson PLC and the Remediation of Methods, European Educational Research Journal 15(1)(2016) 34-53.
[9] C. Troussas, A. Krouska, M. Virvou, Using a multi-module model for learning analytics to predict learners' cognitive states and provide tailored learning pathways and assessment, in: M. Virvou, E. Alepis, G. Tsihrintzis, L. Jain (Eds.), Machine Learning Paradigms: Intelligent Systems Reference Library, Cham: Springer, 158, 2020 (11-29).
[10] Y. Guo, Y. Yang, Construction of the Learner Portrait Based on Online Learning Behavior, Advances in Education 13(2)(2023) 848-857.
[11] M. Khalil, S. Slade, P. Prinsloo, Learning analytics in support of inclusiveness and disabled students: a systematic review, Journal of Computing in Higher Education, 2023, https://doi.org/10.1007/s12528-023-09363-4.
[12] G. Siemens, Learning Analytics: The Emergence of a Discipline, American Behavioral Scientist 57(10)(2013) 1380-1400.
[13] N. Yan, O.T. Au, Online learning behavior analysis based on machine learning, Asian Association of Open Universities Journal 14(2)(2019) 97-106.
[14] T.L. Quy, G. Friege, E. Ntoutsi, A review of clustering models in educational data science towards fairness-aware learning. <https://arxiv.org/ftp/arxiv/papers/2301/2301.03421.pdf>, 2023 (accessed 04.22.23).
[15] A. Dutt, S. Aghabozrgi, M.A.B. Ismail, H. Mahroeian, Clustering Algorithms Applied in Educational Data Mining, International Journal of Information and Electronics Engineering 5(2)(2015) 112-116.
[16] S.L.S. López, R.P.D. Redondo, A.F. Vilas, Discovering knowledge from student interactions: clustering vs classification, in: Proc. TEEM 2017: Proceedings of the 5th International Conference on Technological Ecosystems for Enhancing Multiculturality, 2017.
[17] S.R. Sobral, C.F. de Oliveira, Clustering Algorithm to Measure Student Assessment Accuracy: A Double Study, Big Data and Cognitive Computing 5(4)(2021) 81.
[18] B.K. Francis, S.S. Babu, Predicting Academic Performance of Students Using a Hybrid Data Mining Approach, Journal of Medical Systems 43(6)(2019) 162.