Journal of Advancement in Computing (JAC)

AN EFFECTIVE DATA MINING APPROACH FOR ASSESSING STUDENT'S SATISFACTION IN ONLINE EDUCATION DURING COVID-19 PANDEMIC

Ansar Riaz , Muhammad Hassan , Sufyan Hameed, Muhammad Adeel Imdad
[1]#Faculty of Computing, Student Inn College Faisalabad
[2]#Faculty of Computing, Kips College Faisalabad
[3]# Faculty of Computing, Govt University Faisalabad (GCUF)
[4]# Faculty of Computing, Comsats University Wah Catt
Email: ansar.riaz16@gmail.com, hassanzahid79013@gmail.com, sufyanh216@gmail.com,adeelm69@gmail.com

DOI: https://doi.org/10.36755/jac.v2i1.59

A B S T R A C T

The advent of the COVID-19 outbreak has caused widespread public-health concerns. As a result of these emergency conditions, several governments have opted to implement lockdowns to reduce social interaction and minimize infection. COVID-19 has significantly impacted Higher Education Organizations (HEOs). Many unorthodox educational methods are proposed to ensure the continuation of the learning system in light of the effects of this pandemic and the necessity for alternative remedies. Online Education (OE), also based on learning together in a synchronous or asynchronous environment by employing various equipment, including mobile devices, Computers, and so forth, for Internet access, was among the options. All education systems are primarily concerned with boosting students' academic achievement to improve the overall standard of teaching. In this regard, Educational Data Mining (EDM) seems to be an expeditiously growing field of research that employs the significance of Data Mining (DM) ideas to assist the education system in determining valuable information just on Student Satisfaction Learning (SSL) with both Online Learning procedure (OL) as during COVID-19. Various approaches have been explored using EDM to forecast students' behaviors to provide the optimum educational settings. As a result, Feature Selection (FS) was commonly used to find one of the most indicates. the status of characteristics with the least cardinality. For COVID-19 to find accuracy result in this research KNN and SVM

A R T I C L E I N F O

Article History:
Received 12/09/2023
Accepted 31/01/2024
Online 31/01/2024
Keywords:
Higher education organizations, educational data mining, feature selection, random forest.

algorithm used by using modified data set from Kaggle. Results showed 79.9% precision of education level wise prediction using KNN, 73.7% precision of devices wise prediction using KNN and 88.5% precision of educational level wise predication using SVM, 73.8% precision of device wise prediction using SVM which is showing that the proposed model is significant.

1. INTRODUCTION

Due to the COVID-19 pandemic, many countries had to suspend in-person classes and adopt alternative methods of education, such as online education (OE). Several scholarly studies have been conducted to assess the impact of COVID-19 specifically on the teaching system [1]. However, further research is needed to understand the effects of this epidemic on the learning process. This emphasizes the importance for Higher Education Institutions (HEIs) to develop comprehensive strategies to meet students' educational standards outside traditional classrooms. To address this issue, [2] explored the role of trust in online educational resources in terms of preparation and various approaches, including mass communication and social interaction programming associated with online education. The researchers also proposed university initiatives to enhance attitudes towards OE, such as providing global education opportunities for both professors and students, enabling them to benefit from online learning (OL). Additionally, it is crucial for colleges and universities to consider potential obstacles to student communication in their decision-making processes after an epidemic [3].

1.2 EDUCATIONAL DATA MINING

Educational Data Mining (EDM) is an emerging field focused on extracting information from educational data. Given the significant expansion of the sector, EDM has the potential to accurately predict student achievement. Through statistical analysis software techniques, it is possible to make accurate predictions, categorize data, establish connections, and group students effectively. This enables the provision of additional information about potential student dropouts and their performance in various events, thereby enhancing the quality of education. EDM has the potential to benefit not only students but also other stakeholders involved in the educational process. Currently, interactive e-learning technologies and methods facilitate the collection and analysis of student data. However, due to the dynamic nature of data, retrieving this type of information can be challenging [4].

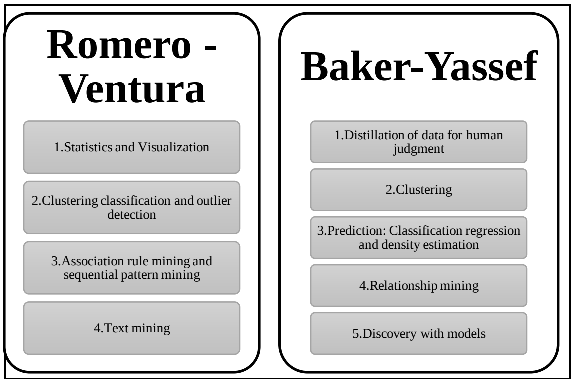

Fig.1 Types of Educational Data Mining

1.3 CLASSIFICATION

Classification is one of the oldest analytical techniques and remains widely recognized and commonly employed in data mining. Bayesian classification, which considers the passing grade as a basis, can be utilized in this context. By applying conditionally Bayesian probability, it becomes possible to assess the likelihood of a student being hired by a global corporation when they achieve an "A" grade. Similarly, Bayesian classification can be employed to determine the probability of selection for students at lower academic levels [5].

1.4 PREDICTION

Regression analysis has been widely utilized as a prevalent prediction method, involving the use of one or more regression models. This technique can be applied to both continuous and categorical variables. The prediction is established by examining the relationship between a known variable and the factor that needs to be predicted [6].

1.5 CLUSTERING

Clustering refers to the process of categorizing a collection of items into groups based on their similarities. Segmentation methods, based on the cluster structure, are typically divided into three distinct categories [7].

2. RECENT WORK

The researcher [8] contributed to the advancement of the field of learning evaluation, which played a central role. It is important to highlight that two organizations, namely the International Educational Data Mining Society and the Society for Learning Analytics Research, initially spearheaded the efforts in establishing reliable assessments in instructional design. During a similar timeframe, the drivers and challenges behind learning evaluation studies were also identified and addressed.

2.1 ONLINE LEARNING DURING THE COVID-19

According to the researchers [9], it was suggested that students who received online education during the COVID-19 pandemic achieved higher academic performance compared to those who did not receive direct instruction from their Higher Education Institution (HEI). The researchers conducted a survey to assess the impact of online learning (OL) on college students during the COVID-19 crisis, analyzing both the positive and negative aspects of OL from the students' perspectives in comparison to traditional learning. Due to the limitations imposed by the COVID-19 outbreak, instructors were compelled to incorporate new teaching methods in order to maintain a satisfactory level of education, which highlighted the excellent adaptability of online learning (OL) in such circumstances [10].

2.2 STUDENT PERFORMANCE ANALYSIS

Identifying the appropriate predictors is essential for achieving successful prediction outcomes, as emphasized by [11]. When predicting student achievement, it is crucial to acknowledge the factors that impact the dissemination of information. Moreover, numerous educational institutions expressed apprehension regarding the decreasing rates of academic achievement and the frequent occurrence of student withdrawals.

2.3 INFLUENCING FACTOR IDENTIFICATION

In their study, the researchers [12] examined the behavioral and educational background of the participants, which encompassed aspects such as behavior, previous assessments, areas for improvement, and prior knowledge of curriculum subjects. They presented a conceptual framework based on the existing literature in their research.

2.4 STUDENT BACKGROUND AND BEHAVIOR

In the evaluation of student characteristics [13], a comprehensive analysis of students' progress was conducted, encompassing various aspects such as behavioral patterns, social interactions, and academic background. Among the extensively discussed topics in this analysis was the impact of gender.

2.5 STUDENT PERFORMANCE PREDICTION

According to [14], the purpose of grading is to assess the extent to which a student learns and applies the information taught in a course. However, accurately estimating the actual score can be challenging. This challenge has motivated additional efforts by scholars in the field of Educational Data Mining (EDM) to develop effective models for describing student achievement. In addition to providing convenience, prediction plays a crucial role in supporting decision-makers, including teachers, in making appropriate and timely interventions.

3. RESEARCH METHODOLOGY

Amidst the COVID-19 lockdown, virtually all higher education institutions transitioned from traditional classroom settings to virtual classrooms. In Pakistan, numerous universities initially relied on the Traditional Education System (TES), which posed challenges for students accustomed to TES. However, when the pandemic emerged, academic institutions made the decision to implement online learning environments to save students' time. Additionally, the assessment of students' problem-solving skills during both TES and Virtual Learning Environments (VLE) was a notable topic of concern, as depicted in the provided image. Lastly, an inquiry was made regarding students' perceptions of VLE [15].

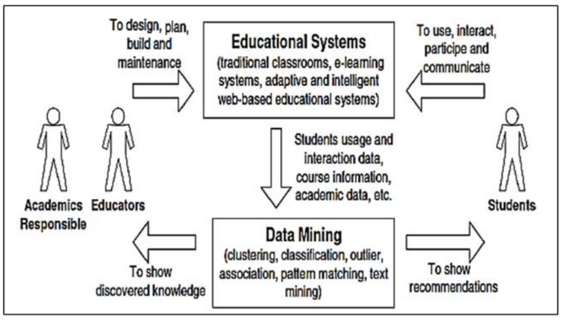

Fig.2 Research Flow

3.1 PRE-PROCESSING

Following the collection of datasets, cleaning is a key step in obtaining correct data and removing ambiguity from data. It entails several phases such as data cleansing, data processing, feature selection, Additionally, in this study, different professors employed various terms to gather input from their students, such as "loss of online connection" or "no internet." Consequently, it was necessary to preprocess the data to address these variations and ensure consistency in the dataset. preprocessed.

- Data Cleaning
- Data Integration
- Data Transformation
- Data Reduction

3.2 PROPOSED MODEL

The primary objective is to develop a prediction model using machine learning (ML) techniques that can achieve the highest possible accuracy in the SSL system. To accomplish this, two classification approaches, namely K-NN and SVM, are employed. These approaches enable the iterative evaluation of the importance, reliability, and quantity of the extracted features throughout the process. The aim is to optimize the model's performance and enhance the accuracy of predictions.

3.3 MACHINE LEARNING TECHNIQUES

A) K-Nearest Neighbors (K-NN)

The K-NN approach is a fundamental supervised machine learning technique commonly used in various machine learning tasks. It is favored for its simplicity in comparison to more complex supervised ML methods, making it widely adopted. K-NN is utilized in various fields within the pattern matching paradigm, including image identification, finance, medicine, and forestry [16].

B) Support Vector Machine (SVM)

SVM, a supervised machine learning technique, is effective in addressing both classification and regression problems. Its primary application lies in solving classification challenges. SVM is highly valued in the field of data mining due to its ability to achieve high classification accuracy while utilizing fewer computational resources.

4. RESULTS AND DISCUSSIONS

This section presents the results of the study. The initial research question explores the impact of virtual classrooms on student achievement. Teachers gathered feedback from students through online forms, and a survey was conducted to assess the challenges faced by students and the outcomes they experienced.

4.1 PROBLEMS FACED BY STUDENTS DURING VIRTUAL LEARNING ENVIRONMENT

The graphs below represent the mean of the issues encountered by BS candidates during VLE.

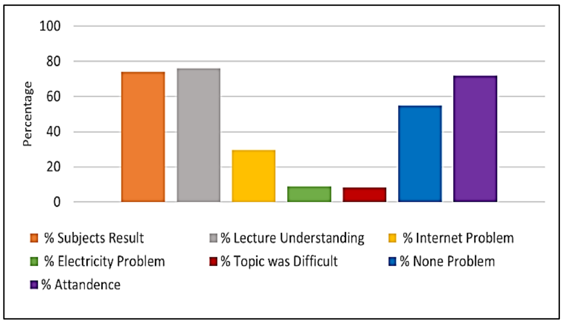

Fig.3 Problems by BS Subject Average during VLE.

The provided chart illustrates the average challenges experienced by BS candidates. The primary outcome indicates that the comprehension of the subject matter was rated at 74%, the readability of the lessons at 76%, internet connectivity issues at 29%, electricity-related challenges at 9%, finding the teaching topics difficult at 8%, and no difficulties reported by 54% of the students. Furthermore, the participation rate during the virtual classroom was recorded at 72%.

4.2 PROBLEMS FACED BY MS DURING VIRTUAL LEARNING ENVIRONMENT

The following charts display the average challenges experienced by MS participants during Virtual Learning Environments (VLE). The findings include an analysis of the difficulties encountered by male and female students, providing insights into their respective challenges throughout the VLE.

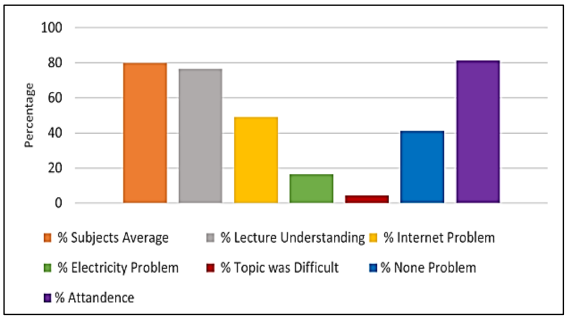

Fig.4 Problems by MS Students Average

The provided chart presents the average challenges faced by MS candidates. The results indicate that forward-thinking abilities were rated at 80%, presentation readability at 76%, internet connectivity issues at 49%, electricity-related difficulties at 16%, finding the presentation topics challenging by 4% of the students, no difficulties reported by 41% of the students, and a participation rate of 81% during the virtual classroom.

4.3 CORRELATION OF UNDERSTANDING OF STUDENTS WITH DIFFERENT FACTORS

Examine the relationship between the students' comprehension and the challenges they encounter inside the VLE.

Table.1 Correlation of Understanding of Students with Different Factors

Sr.	Subject Name	Internet Problem	Electricity Problem	Topic Difficulty	No problem	Attendance
1	Discrete structure	-0.66	-0.77	0.032	0.71	0.24
2	Internet of Things	0.27	-0.76	-0.22	-0.28	0.21
3	Introduction to Computing	-0.64	-0.47	-0.85	0.87	-0.14
4	Blockchain Technology	-0.52	-0.49	-0.12	0.66	-0.05
5	Operating System	-0.58	-0.47	-0.28	0.78	0.41
6	Machine	0.51	-0.05	0.46	-0.55	-0.15

7	Learning Average of all subject	-0.57	-0.48	-0.45	0.77	-0.21

4.4 RESULTS FOR PREDICTIONS

In this section, the author employed a range of machine learning techniques, including SVM and KNN, to delve deeper into the concepts being explored. The author utilized SVM and KNN as part of their investigation and analysis.

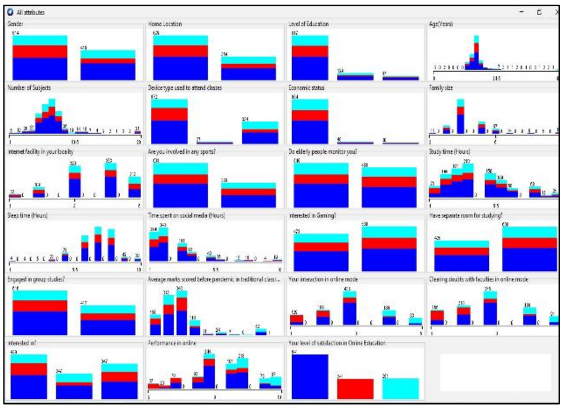

Fig.5 Visualizing All Attributes

The applied SVM and KNN on each attribute to explore and predict. Proposed attributes are (Gender Wise, Education Level, Device Wise, Economic Status Wise, Internet Facility-based, Average Study Time, Study Environment-based, Before Online Grades, Interaction with Online Studies, Performance During Online

Education, Satisfaction Level with Online Learning)

4.4.1 RESULTS USING ALGORITHM KNN AND SVM

Table.2: Economic Status Wise Prediction using KNN , SVM

Algorithm	Parameter	Class		
		Middle	Poor	Rich
KNN	Precision	0.929	0.143	0.083
	Recall	0.929	0.111	0.125
	F-Measure	0.929	0.125	0.100
SVM	Precision	0.924	---	---
	Recall	1.000	---	---
	F-Measure	0.960	---	---

.

Table .3 Education Level Wise Prediction using KNN , SVM

Algorithm	Parameter	Class		
		Under graduate	Post graduate	School
KNN	Precision	0.799	0.126	0.194
	Recall	0.825	0.174	0.194
	F-Measure	0.811	0.193	0.194
SVM	Precision	0.808	0.885	---
	Recall	0.996	0.178	0.000
	F-Measure	0.893	0.297	---

Table .4 Device Wise Prediction using KNN , SVM

Algorithm	Parameter	Class		
		Laptop	Desktop	Mobile
KNN	Precision	0.737	0.000	0.506
	Recall	0.816	0.000	0.425
	F-Measure	0.775	0.000	0.462
SVM	Precision	0.738	---	0.717
	Recall	0.924	0.000	0.410
	F-Measure	0.820	---	0.522

Table.5 Gender Wise Prediction using KNN , SVM

Algorithm	Parameter	Class	
		Male	Female
KNN	Precision	0.676	0.449
	Recall	0.655	0.473
	F-Measure	0.665	0.461
SVM	Precision	0.674	0.630
	Recall	0.837	0.406
	F-Measure	0.747	0.493

Table .6 Satisfaction Level with Online Learning using KNN , SVM

Algorithm	Parameter	Class		
		Average	Bad	Good
KNN	Precision	0.543	0.403	0.352
	Recall	0.628	0.349	0.284
	F-Measure	0.582	0.374	0.314
SVM	Precision	0.593	0.659	0.736
	Recall	0.891	0.353	0.267
	F-Measure	0.712	0.459	0.392

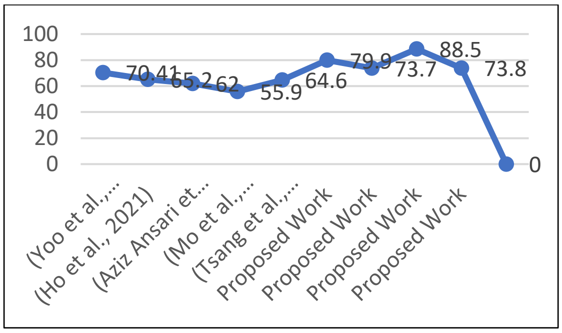

Fig .6 Comparative Analysis Cart

The Figure demonstrates that in 2021 [17] algorithm result is 65.2%. The other compared method that used logistic regression in the year 2022 [18], the algorithm result was 70.41%. A other compared method that used a computer-based cross-sectional study in the year 2021 [19], the algorithm result is 62%. The other compared method that used Smart PLS in the year 2021 [20], the algorithm result was 55.9%. A other compared method that used the Structural model in the year 2021 [21-24] result was 64.6%. Thus, the proposed method gives a distinction in achieving the results.

5. CONCLUSION

The approach involved the collection and analysis of various data sources, including students' feedback and performance, to identify patterns and factors affecting student satisfaction.

The results showed that the approach was effective in identifying significant factors that influence student satisfaction, such as course design, instructor support, and course materials. These insights can help educators and administrators to make data-driven decisions to improve the quality of online education and enhance student satisfaction. Furthermore, the approach can be applied in other educational contexts beyond the COVID-19 pandemic and can help to improve online learning experiences in the future.

In summary, this study highlights the importance of using data mining techniques to assess student satisfaction in online education. It provides valuable insights into the factors that affect student satisfaction and can guide educators and administrators in making informed decisions to improve the quality of online education.

RECOMMENDATIONS

1. Continuously collect and analyze student feedback and performance data to identify patterns and trends in student satisfaction. This will help educators and administrators to address issues and improve the quality of online education.

2. Implement effective communication strategies between instructors and students to enhance instructor support, course materials, and overall course design. Regular check-ins, Q&A sessions, and feedback mechanisms can help to improve the online learning experience for students.

3. Use data mining techniques to analyze student performance data to identify areas of weakness and strength. This can help instructors to tailor their teaching methods to address student needs and improve learning outcomes.

4. Collaborate with other educational institutions to share best practices and benchmark performance against industry standards. This will help to improve the quality of online education across the board.

5. Provide ongoing training and support to instructors to help them develop the skills and knowledge needed to deliver effective online courses. This can include training in instructional design, communication strategies, and online learning technologies.

These recommendations can help to improve the quality of online education during the COVID-19 pandemic and beyond. By using data mining techniques to assess student satisfaction and identify areas for improvement, educators and administrators can enhance the online learning experience for students and help them to achieve their educational goals.

REFERENCES

[1] Alam, T., Gupta, R., Qamar, S., & , A. (2022). Recent applications of Artificial Intelligence for Sustainable Development in smart cities. In Recent Innovations in Artificial Intelligence and Smart Applications (pp. 135-154). Cham: Springer International Publishing.

[2] Sebai, D., & Shah, A. U. (2023). Semantic-oriented learning-based image compression by Only-Train-Once quantized autoencoders. Signal, Image and Video Processing, 17(1), 285-293.

[3] Alam, T., Ullah, A., & Benaida, M. (2023). Deep reinforcement learning approach for computation offloading in blockchain-enabled communications systems. Journal of Ambient Intelligence and Humanized Computing, 14(8), 9959-9972.

[4] Ouhame, S., Hadi, Y., & , A. (2021). An efficient forecasting approach for resource utilization in cloud data center using CNN-LSTM model. Neural Computing and Applications, 33, 10043-10055.

[5] Aznaoui, H., A., Raghay, S., Aziz, L., & Khan, M. H. (2021). An efficient GAF routing protocol using an optimized weighted sum model in WSN. Indonesian Journal of Electrical Engineering and Computer Science, 22(1), 396-406.

[6] Ullah, A., & Nawi, N. M. (2021). An improved in tasks allocation system for virtual machines in cloud computing using HBAC algorithm. Journal of Ambient Intelligence and Humanized Computing, 1-14.

[7] Ullah, A., Nawi, N. M., & Ouhame, S. (2022). Recent advancement in VM task allocation system for cloud computing: review from 2015 to2021. Artificial Intelligence Review, 1-45.

[8] Hussein, E., Daoud, S., Alrabaiah, H., & Badawi, R. (2020). Exploring undergraduate students' attitudes towards emergency online learning during COVID-19: A case from the UAE. Children and Youth Services Review, 119(8), 105-699.

[9] [Clark, A. E., Nong, H., Zhu, H., & Zhu, R. (2021). Compensating for academic loss: Online learning and student performance during the COVID-19 pandemic. China Economic Review, 68(8), 1-14.

[10] Ali, M. M., Paul, B. K., Ahmed, K., Bui, F. M., Quinn, J. M., & Moni, M. A. (2021). Heart disease prediction using supervised machine learning algorithms: Performance analysis and comparison. Computers in Biology and Medicine, 136(10), 104-672.

[11] Kovacic, Z. (2010). Early prediction of student success: mining students enrolment data. Proceedings of the 2010 In Informing Science Information Technology Education Joint Conference, 19th & 24th June, Italy (pp. 647–665). InSITE.

[12] Kardan, A. A., Sadeghi, H., Ghidary, S. S., & Sani, M. R. F. (2013). Prediction of student course selection in online higher education institutes using neural network. Computers and Education, 65(13), 1–11.

[13] Bovo, A., Sanchez, S., Héguy, O., & Duthen, Y. (2013). Clustering moodle data as a tool for profiling students. Second international conference on E-Learning and E-Technologies in

[14] education, 23th & 25th, September Poland (pp. 121-126). IEEE.

[15] Chrysafiadi, K., & Virvou, M. (2013). Student modeling approaches: A literature review for the last decade. Expert Systems with Applications, 40(11), 4715–4729.

[16] Clark, A. E., Nong, H., Zhu, H., & Zhu, R. (2021). Compensating for academic loss: Online learning and student performance during the COVID-19 pandemic. China Economic Review, 68(8), 1-14.

[17] Maqableh, M., Jaradat, M., & Azzam, A. (2021). Exploring the determinants of students' academic performance at university level: The

mediating role of internet usage continuance intention. Education and Information Technologies, 26(4), 4003–4025.
[18] Helal, S., Li, J., Liu, L., Ebrahimie, E., Dawson, S., & Murray, D. J. (2019). Identifying key factors of student academic performance by subgroup discovery. International Journal of Data Science and Analytics, 7(3), 227–245.
[19] Zollanvari, A., Kizilirmak, R. C., Kho, Y. H., & Hernandez-Torrano, D. (2017). Predicting students' gpa and developing intervention strategies based on self-regulatory learning behaviors. IEEE Access, 5(17), 23792–23802.
[20] Ahmed, S., Paul, R., & Hoque, A. S. M. L. (2014). Knowledge discovery from academic data using association rule mining. International Conference on Computer and Information Technology, 22th & 23th December, Bangladesh (pp. 314–319). IEEE.
[21] Quille, K., & Bergin, S. (2018). Programming: Predicting student success early in CS1. A re-validation and replication study. Annual Conference on Innovation and Technology in Computer Science Education, 2nd July, Cyprus (pp. 15–20). ITiCSE.
[22] Phungsuk, R., Viriyavejakul, C., & Ratanaolarn, T. (2017). Development of a problem-based learning model via a virtual learning environment. Kasetsart .
[23] Hasan, R., Hossain, M. M., & Khan, R. (2015, March). Aura: An iot based cloud infrastructure for localized mobile computation outsourcing. In 2015 3rd IEEE International Conference on Mobile Cloud Computing, Services, and Engineering (pp. 183-188). IEEE.
[24] Haghnegahdar, L., Joshi, S. S., & Dahotre, N. B. (2022). From IoT-based cloud manufacturing approach to intelligent additive manufacturing: Industrial Internet of Things—An overview. The International Journal of Advanced Manufacturing Technology, 1-18.