THE FACTOR AFFECTING STUDENT'S PERFORMANCE OF E-LEARNING ENVIRONMENT USING MACHINE LEARNING ALGORITHM

Devanshu Kumar[1], Md. Alimul Haque[2], Khushboo Mishra[3], Farheen Islam[4],
[5]Binay Kumar Mishra
[1,2,5]Department of Computer Science, Veer Kunwar Singh University, Ara- 802301, India
[3]Department of Physics, Veer Kunwar Singh University, Ara- 802301, India
[4]Department of Education, Patna Women's College, Patna, India
Corresponding author: *shadvksu@gmail.com
https://orcid.org/0000-0002-0744-0784

Abstract:
The Covid-19 pandemic is affecting many aspects of society, especially university educational programs worldwide. As a result, online learning is an effective strategy that is adopted by many educational institutions nowadays. However, not all training institutions have the necessary environment, assets, and ability to conduct effective online learning, particularly in poor nations with resource constraints. As a result, many institutions are struggling to build traditional courses or E-Learning in limited condition while still meeting students' demands. To overcome this limitation, we present a technique for assessing the impact of these elements on the e-learning system. Then, utilizing data from students who have participated in the program, this is an issue of explaining the significance and prioritizing construction investments for every component based on the K-means clustering algorithm. The purpose of this paper is to investigate the relationship between the students' responses to e-learning platforms and their performance in terms of various skill levels with the help of K-Means clustering algorithms. The clustering findings demonstrate that individuals with greater levels of involvement outperform those with intermediate or lower levels of engagement.

Keywords: **P**erformance level, E-Learning Technology, Machine Learning, K-means Clustering, Blended Learning

1. Introduction

The Covid-19 outbreak is quickly expanding over the globe and shows no certain sign of slowing down. Furthermore, the outbreak has impacted all areas of business and politics, as well as the lifestyle of all people globally. This isolation necessary to stop the transmission of the virus has resulted in the formation everyday new lifestyle of people of society via the World Wide Web. Schooling is among the most impacted areas since it is essential for many people throughout every nation. Students' capacity to attend classes and participate in group activities is restricted, thus education, researching, and exchange activities are mostly carried out through e-learning or blended learning techniques(Mseleku, 2020).

Students are supposed to undergo their studies online as educational institutions are closed due to safety measures in the current pandemic situation(Almaiah et al., 2020). The students must grasp the subjects with clear understanding, which is possible by enhancing the quality of education. The fundamental reason for the decline in the quality of education is that students cannot comprehend and understand the subject(Maatuk et al., 2021).

The combination of e-learning and traditional training has provided universities with an effective method of training. This method allows them to provide their students with an online platform that enables them to complete their tasks without being physically present. Despite the advantages of e-learning, it still has certain limitations. Aside from having the necessary skills and expertise in technology, the system also needs to be secure and synchronous.

E-learning is a phrase used in the education industry to describe the use of information technology tools and communication methods. In both local and institutional situations, e-learning has a favorable influence on the education industry. E-learning offers a variety of instruments for students' management as well as instructional needs at the institutional level. The administrative aspects, such as student enrolment, and the instructional aspects, such as course content and online tests, are both important. Learning activities are conducted at the local level instead of in a global setting. This method of education is very important to sustaining the continuous development of the economy.

As a consequence of various investigations, cloud computing has been presented as a solution for higher resource utilization than traditional servers. Because of its flexibility, speed, scalability, and flexibility, cloud computing has been offered as a unique way for enabling IT solutions(Kuttattu et al., 2019). When using an e-learning system on a cloud platform, a variety of benefits may be acquired; the following list summarizes a few of these advantages: the main advantage is that the e-learning system's performance will be improved since the system's services will be put on the cloud environment. By deploying the system's services on a cloud computing platform, you may save money on things like maintenance and management(Kuttattu et al., 2019). Another advantage is that the system's services will always be available to the system's users. Students and instructors can access the services using a variety of devices from various locations.

The behavior of a student in class is influenced by a number of things. Compared to face-to-face learning, online learning is significantly different. Because of the nature and setting of learning, students are finding it challenging to cope with online learning. While they are at home, they also have other responsibilities to attend to.

Instructors may generally examine their students immediately during face-to-face learning and ensure that they stay attentive in class. Instructors in online learning, on the other hand, can only assess pupils digitally. It is not necessary for pupils to be able to discussion in class. Both instructor and learner are having difficulty adjusting to their new surroundings. This condition is particularly troubling for kids, as it has an influence on their academic performance. Online learning, in comparison to traditional learning, necessitates a higher level determination to do their best in the virtual classroom(Pustika, 2020).

This study tries to explore the relationship among a participant's reply on digital platforms as well as their achievement in classroom by studying their behavior while utilising an e-learning system. The study attempts to categories groups that demand involvement and interest in the teaching and learning activities using clustering method by the Machine Learning algorithm such as K-means algorithm(Aggarwal & Sharma, 2019).

2. Materials and Methods

2.1 Dataset

The dataset applied in this research is "E-Learning Student Reactions: Students Reactions and Posts from E-Learning," which is openly accessible on Kaggle(Abdo et al., 2021).These dataset was gathered over the course of four months in a university's algorithm introduction session. It is made up of 16 variables and 71 occurrences, all of which were gathered as quantitative data. Every row represents a metric of a single participant, as well as the types of reactions from other students and the instructor's skill grade.

2.2 Data Clustering

Data clustering aids in the discovery of underlying patterns and the classification of these patterns into distinct groupings known as clusters. The K-means clustering technique is used in this example. When there are a large number of variables, the k-means approach is computationally quicker than alternative clustering techniques(Yusuf et al., 2020). The first stage in this approach is to assign k centroids to each other at random. The Euclidean distance between each data point and the centroid is determined once the centroids have been identified(Cioruța et al., 2021). The centroid nearest to each data point is then assigned. After that, the centroids are repositioned in order to achieve the best results. These steps are performed repeatedly until the centroids do not change. We get fifty one clusters after using the k-means technique(Ahmed et al., 2020)(Abdo et al., 2021). A cluster number will be assigned to each and every piece of data. After data clustering, we add this cluster number as a feature to the dataset, bringing the total number of features to sixteen. We can divide the dataset into test and train sets once the cluster number has been added to it. In this example, we'll divide the data set in half, with one serving as the test set and the other as the train set(Kuttattu et al., 2019).

The Elbow approach was utilized to determine the optimal k during the construction of the K-Means algorithm. The Elbow technique computes the sum of squared errors (SSE) for each of the k variables. The number of clusters employed is three, according to the Elbow technique(Abdo et al., 2021). The square of the distance between the data point and the cluster center is used to compute the SSE (Nainggolan et al., 2019).

2.3. Implementation

The complete dataset was utilized for training after pre-processing such as missing value treatment, data inconsistencies, and so on. The clustering technique was started with three clusters using the Elbow approach. Students, their total responses, and the overall grade of their talents were used to create the clusters. The first grouping wasn't considered adequate. MinMaxScaler was used to normalize the feature scale in order to improve clustering. The scaling aided in the creation of a more accurate clustering picture. The findings will be addressed in the following sections.

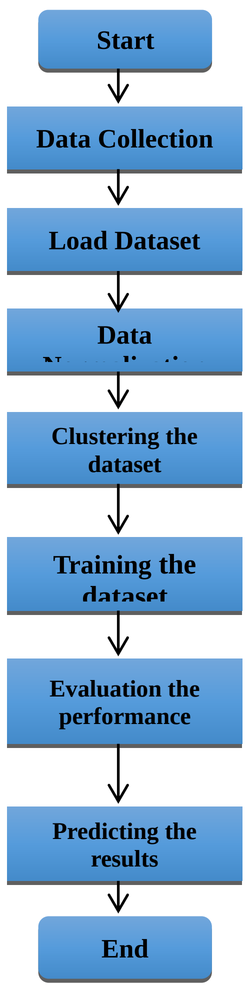

Fig. 1: Steps involved

3. Results & Discussion

The data findings are presented in this part, with a focus on the link between the three key elements of student behavior, peer reaction, and skill grades. The average abilities grade ranges from 0 to 9 on the histogram in Fig 2. The majority of the students in the class have a skills grade between 0 and 1 and an average of 8.

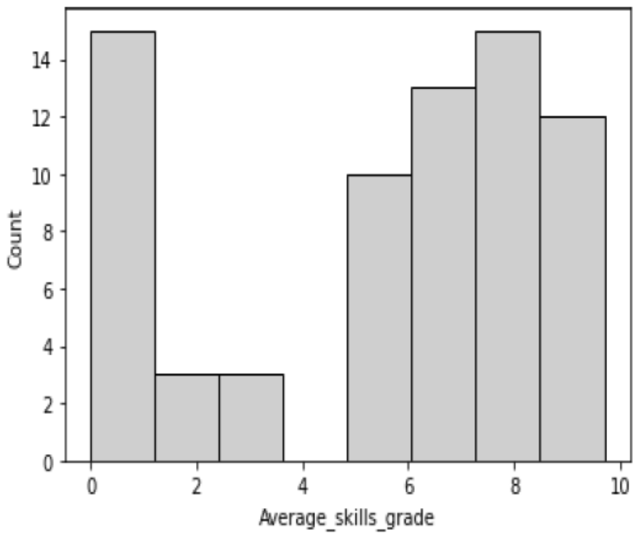

Fig. 2: Average grade for skills

The scatterplot in Fig 3 shows the correlation between a student's final grade and the total number of posts they made in the online class. The scatter plot is positive, linear, and somewhat skewed. The upward trend demonstrates that the overall number of posts made by students reflects strong student achievement. As a result, the bigger the number of posts generated by students, the higher their average class skill grade.

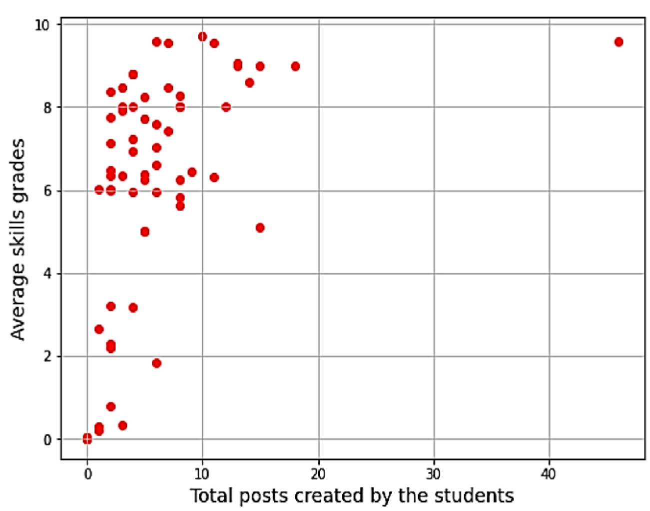

Fig. 3 Students posts against average skills grade

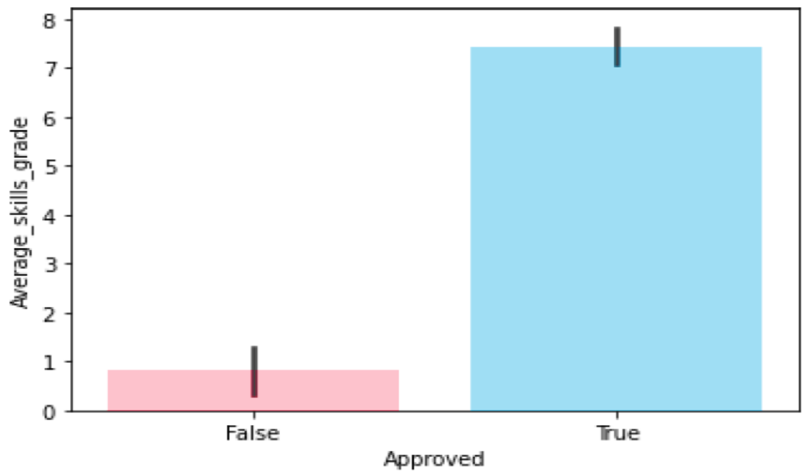

Fig. 4: Students average and instructor approval

The student's average grade in class is plotted against the student's approbation in class in Figure 4. It can be seen that the majority of the pupils in the class that are accepted have a high average grade. If you get good grades, it appears that you will get more approbation in class.

Figures 5 and 6 show the total response and time spent online, as well as the students' overall final grade. Between the values of the variables, there is an increasing trend. Based on this pattern, it can be stated that total response and online time help students perform better. Students who participate actively in class likely to achieve higher grades.

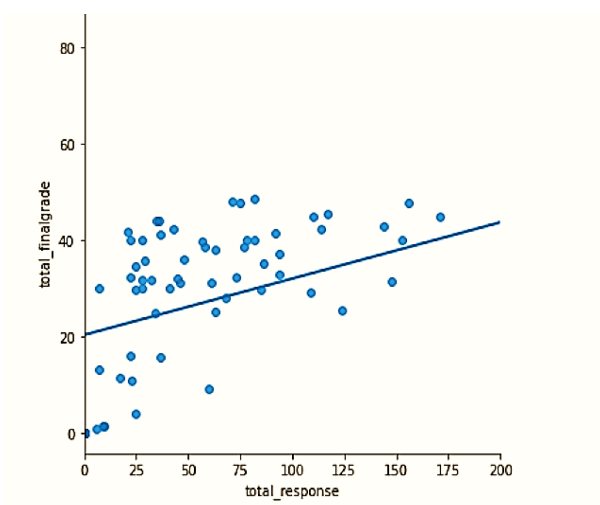

Fig. 5: Total response against total grade of skills

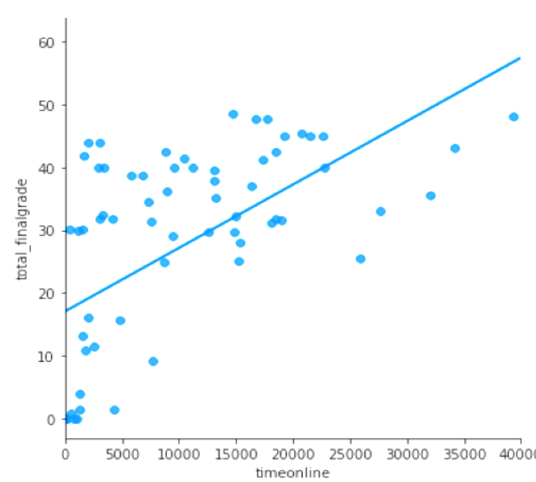

Fig. 6: Time spent against total grade of skills

The Elbow technique is depicted in Fig 7 in the plot that we used to calculate the ideal k. We have three groups since the most elbow-shaped form from the plot above falls on the scale of three.

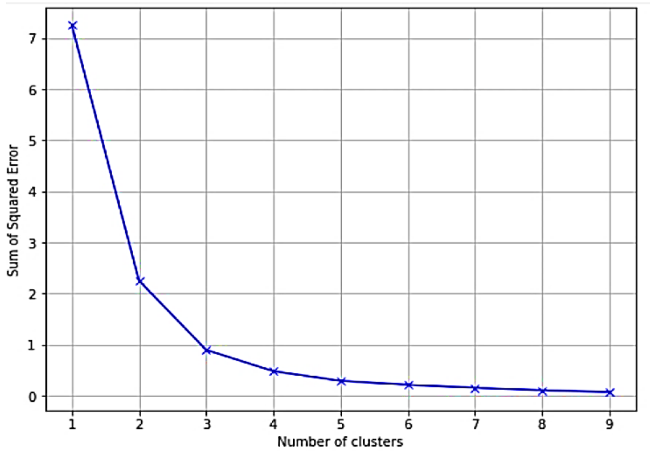

Fig. 7 Elbow method for optimal k

The effect of the K-Means method on the total response and overall skills grade is represented in Fig 8 via a scatter plot. There should've been three clusters based on the Elbow technique. However, without optimizing and normalizing the clustering model, the results show that the majority of students in clusters 0 and 2 are in cluster 1. Cluster 1 has only one data point, which is separated from the other two clusters by a significant amount of time. As a result of applying the MinMaxScaler to normalize the data, the three clusters can be clearly seen in Figure 9.

Low response, medium response, and high response are the three clusters, with low response corresponding to cluster 1 (green), medium response to cluster 2 (red), and high reaction corresponding to cluster 3 (orange).

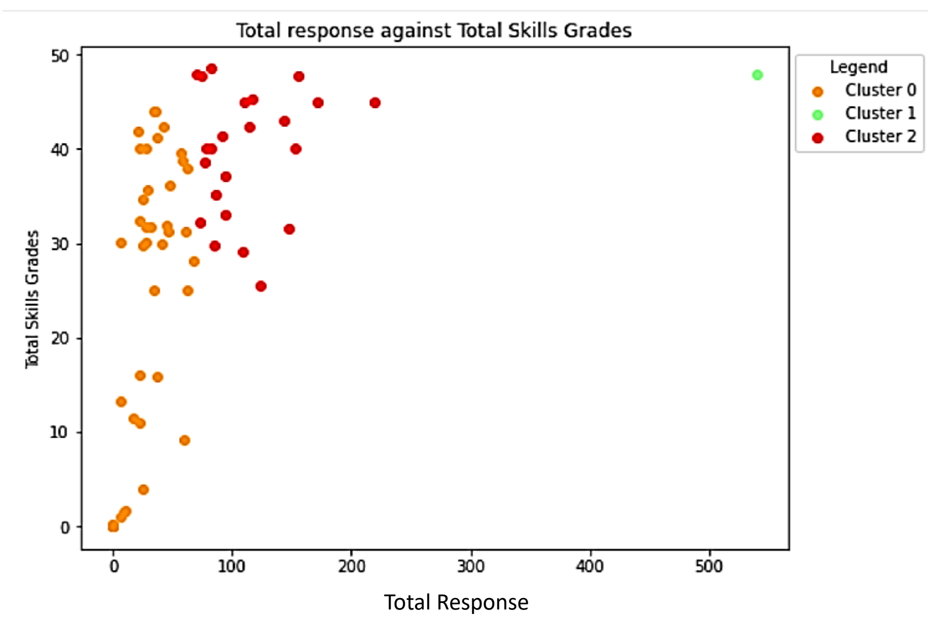

Fig. 8: K-Means clustering

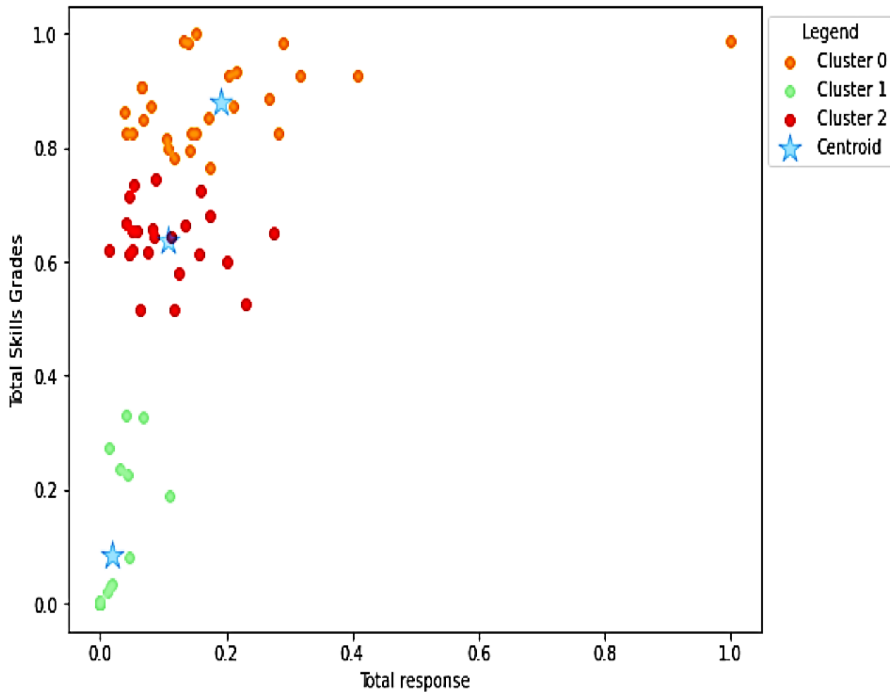

Fig. 9: Normalized K-Means clustering

Cluster 0 is made up of students who frequently respond on the e-learning platform and have a higher overall grade for skills than the other two groups. Cluster 1 comprises of students who seldom react and have a lower overall grade for skills, whereas cluster 2 consists of kids who show a medium ground in delivering reactions and have a median total score for their abilities. Figure 12 also shows the centroid of each cluster. Because the three clusters are unique, the k-value of 3 is able to track and forecast performance based on the data's underlying properties.

4. Conclusion

The study looked at the link between student involvement and performance on e-learning systems. A student who is actively participating in the debate is more interested in the subject at hand. The student's contribution to the topic of study extends beyond passive learning, where the student tends to give their analysis and knowledge. This allows for improved comprehension because the discussion subject is not just drawn from the lecture materials, and it also increases the likelihood of learning even more complex content, particularly in detail. In both online and offline modes of teaching and learning, such measurements and student grouping are necessary on a regular basis. This will allow teachers to focus on groups with lower levels of engagement, determine the cause, and change the lesson and materials as needed.

In order to further generalize the results, such investigations must be duplicated with a dataset with bigger examples. For deeper insights, the data should contain other parameters like attendance, moral behavior, and others. In addition, a variety of clustering methods should be utilized to create a broad picture of the issue behavior.Figure 1 shows how much time students spend on various types of learning activities on a typical week (in hours).Future studies may be focused on various (IoT)Internet of things and implementation strategies, as well as instances of real employment and outcomes. Suitable structures for e-Learning universities ready to accept IoT must be researched for future projects as well(Haque et al., 2021)(Alimul Haque M., Haque S., Rahman M., Kumar K., 2022)(Almrezeq et al., 2022).

References:

Abdo, A. M., Rasid, N. M. A., Badli, N. A. H. M., Sulaiman, S. N. A., Wani, S., & Zainol, Z. (2021). Student's Performance Based on E-Learning Platform Behaviour using Clustering Techniques. *International Journal on Perceptive and Cognitive Computing*, *7*(1), 72–78.

Aggarwal, D., & Sharma, D. (2019). Application of clustering for student result analysis. *Int J Recent Technol Eng*, *7*(6), 50–53.

Ahmed, M., Seraj, R., & Islam, S. M. S. (2020). The k-means algorithm: A comprehensive survey and performance evaluation. *Electronics*, *9*(8), 1295.

Alimul Haque M., Haque S., Rahman M., Kumar K., Z. S. (2022). Potential Applications of the Internet of Things in Sustainable Rural Development in India. *Proceedings of Third International Conference on Sustainable Computing*, 455–467. https://doi.org/10.1007/978-981-16-4538-9_45

Almaiah, M. A., Al-Khasawneh, A., & Althunibat, A. (2020). Exploring the critical challenges and factors influencing the E-learning system usage during COVID-19 pandemic. *Education and Information Technologies*, *25*(6), 5261–5280.

Almrezeq, N., Haque, M. A., Haque, S., & El-Aziz, A. A. A. (2022). Device Access Control and Key Exchange (DACK) Protocol for Internet of Things. *International Journal of Cloud Applications and Computing*, *12*(1), 1–14. https://doi.org/10.4018/IJCAC.297103

Cioruța, B.-V., Lauran, M., Coman, M., Pop, A. L., & Lauran, A. (2021). About the Benefits of Adopting E-Learning in the Current Romanian Educational System. *Asian J. Educ. Soc. Stud*, *15*(3), 1–13.

Haque, S., Zeba, S., Alimul Haque, M., Kumar, K., & Ali Basha, M. P. (2021). An IoT model for securing examinations from malpractices. *Materials Today: Proceedings*. https://doi.org/10.1016/j.matpr.2021.03.413

Kuttattu, A. S., Gokul, G. S., Prasad, H., Murali, J., & Nair, L. S. (2019). Analysing the learning style of an individual and suggesting field of study using Machine Learning techniques. *2019 International Conference on Communication and Electronics Systems (ICCES)*, 1671–1675.

Maatuk, A. M., Elberkawi, E. K., Aljawarneh, S., Rashaideh, H., & Alharbi, H. (2021). The COVID-19 pandemic and E-learning: challenges and opportunities from the perspective of students and instructors. *Journal of Computing in Higher Education*, 1–18.

Mseleku, Z. (2020). A literature review of E-learning and E-teaching in the era of Covid-19 pandemic. *SAGE*, *57*(52), 6.

Nainggolan, R., Perangin-angin, R., Simarmata, E., & Tarigan, A. F. (2019). Improved the performance of the K-means cluster using the sum of squared error (SSE) optimized by using the Elbow method. *Journal of Physics: Conference Series*, *1361*(1), 12015.

Pustika, R. (2020). Future English teachers' perspective towards the implementation of e-learning in Covid-19 pandemic era. *Journal of English Language Teaching and Linguistics*, *5*(3), 383–391.

Yusuf, B., Walters, L. M., & Sailin, S. N. (2020). Restructuring Educational Institutions for Growth in the Fourth Industrial Revolution (4IR): A Systematic Review. *International Journal of Emerging Technologies in Learning*, *15*(3).