IJASST International Journal of Applied Sciences and Smart Technologies

Volume 07
Issue 01

p-ISSN 2655-8564, e-ISSN 2685-943, pp. 1-14, 2025

Cluster Change Analysis to Assess the Effectiveness of Speaking Skill Techniques Using Machine Learning

Nunik Herani Wulandari[1], Vega Purwayoga [2*]

[1]*University of Muhammadiyah Tasikmalaya, Indonesia*
[2] *Department of Informatics, Faculty of Engineering, Siliwangi University, Indonesia*
[*]*Corresponding Author: vega.purwayoga@unsil.ac.id*

(Received 30-08-2024; Revised 22-09-2024; Accepted 07-10-2024)

Abstract

Students have different characteristics in the learning process, especially in developing speaking skills. So it requires the right learning method. This study aims to compare effective teaching methods for speaking using machine learning. The classes used in this study consisted of three classes: conventional, vlog project, and picture series. The data used were students' pre-test and post-test scores. The machine learning algorithm used is K-Means. K-Means clusters the pre-test and post-test data. The results of K-Means clustering on the pre-test and post-test identified the differences in student groups between the pre-test and post-test. Students who experienced the most cluster movement were those in the vlog project class, the conventional class, and the picture series class.

Keywords: Clustering, K-Means, Machine Learning, Speaking Skill

1 Introduction

Speaking is one way to communicate with someone. The main components of speaking include fluency, comprehension, grammar, vocabulary, and pronunciation [1]. Speaking skills do not appear instantly, so regular practice is needed. One of the places to practice speaking skills is at school with the help of teachers [2]. Intensive learning can improve speaking skills, especially in English lessons [3].

Based on research [4], students need to develop skills in speaking English. The earlier students learn English, the easier it is for them to understand English. Learning English is an investment for a bright future. Speaking English allows us to easily access and obtain information because Most people write information in English.

This work is licensed under a Creative Commons Attribution 4.0 International License

Some effective teaching techniques for learning to speak include vlog projects and picture series [5]. The picture series technique allows educators to see the development of students' speaking skills. The picture series learning model positively impacts speaking ability [6]. There is a significant difference between classes that use conventional learning methods and classes that use the picture series learning method [7]. Vlog projects can improve speaking skills by an average of 27%. Vlog projects are the preferred learning method, and a student's level of desire for the method is 89% [8]. Like previous research, research [9] also revealed that vlog projects are an effective learning method to improve students' speaking skills.

Some previous studies have discussed the effectiveness of picture series and vlog project learning methods to improve students' ability to speak English [6]. Few discuss which comparison is more effective than the picture series and vlog project. Research that compares picture series and vlog project learning methods is research [5]. The study's results [5] stated that the vlog project was more effective than the picture series. Comparing effective learning methods in speaking skills can be done with machine learning techniques [10]. Research [11] applied a machine learning algorithm, K-Means, to evaluate effective learning methods. K-Means effectively cluster and evaluate student performance [12]. This research continues the research conducted by previous research [5]. Since K-Means works effectively and few studies compare the learning methods of picture series and vlog projects, this research aims to group students based on pre-learning and post-learning results. This research has contributed in terms of verifying cluster changes that have not been carried out by previous studies. Grouping children is expected to determine the changes or consistency of children's groups when Educators use the picture series or vlog project learning method.

2 Material and Methods

Study Area and Research Data

This research was conducted on the eighth-grade students at MTs Negeri 13 Ciamis located at Jalan Cipancur No. 6 Sirnabaya, Rajadesa, Ciamis Regency, West Java, in the even semester of 2020 / 2021. The researcher conducted research from January to April 2021. This research has several stages as presented in Figure 1.

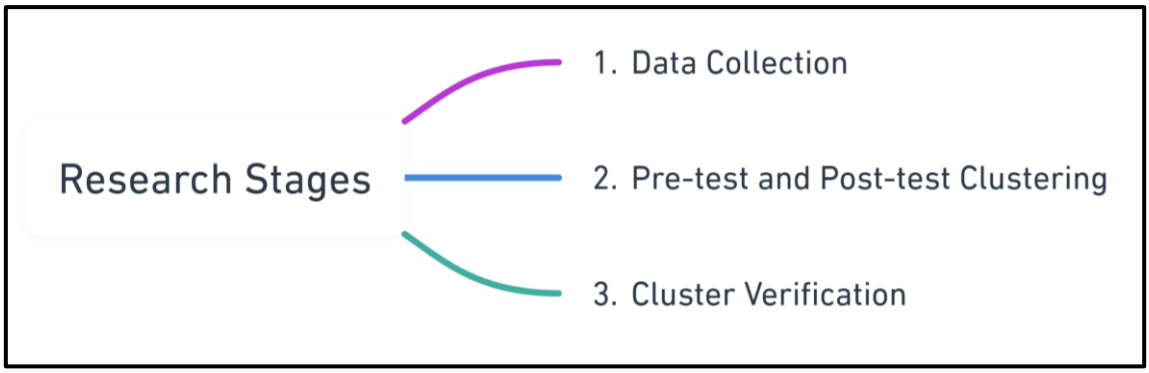


Figure 1. Research Stages

Data Collection

Data collection in the study was to collect pre-test and post-test scores of students from three classes, namely, conventional, vlog project, and picture series classes. The pre-test data collection process was carried out before the learning process, while the students took the post-test after the learning process. The learning process is carried out based on each class. In conventional classes, the English language subject teachers at the school carry it out.

Pre-test and Post-test Clustering

The K-Means cluster is applied using R programming. R Programming is a high-level language easily applied to data mining techniques [13]. K-Means is a popular clustering algorithm. The way K-Means works is to partition data into several groups based on the characteristics of the data. The first object in a cluster can be used as the cluster center point. The K-Means algorithm will repeat the following steps until the center point of a cluster does not change [14]. The stages of the K-Means algorithm can be seen in Figure 2.

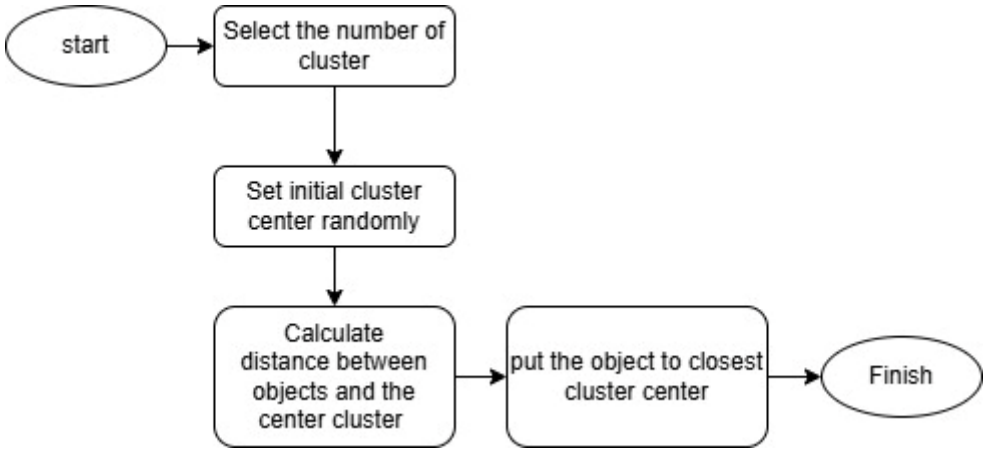


Figure 2. The stages of the K-Means algorithm [15][16]

Cluster Verification

Cluster verification is conducted to determine whether there is a shift or change in a student group after the treatment of each learning technique.

3 Results and Discussions

Data Collection

Data was collected by taking pre-test and post-test data from all classes. The classes consisted of a vlog project class, a picture series class, and a class that applied conventional techniques. The number of students in a class has been presented in Figure 3. The results of the pre-test and post-test data collection from the vlog project, picture series, and conventional classes are presented in Table 1, Table 2, and Table 3, respectively.



Figure 3. Number of students

Figure 2 is a histogram that explains the number of students in each class. The number of students in the conventional class consists of 20 students, the picture series class is 18, and the vlog project class consists of 19 students.

Table 1. Pre-test and post-test of vlog project class

No.	Student	Pre-test	Post-test
1	V1	44	66
2	V2	46	67
3	V3	46	67
4	V4	46	67
5	V5	44	65
6	V6	48	70
7	V7	45	60
8	V8	50	71
9	V9	52	72
10	V10	48	71
11	V11	45	66
12	V12	45	70

13	V13	48	70
14	V14	44	66
15	V15	50	71
16	V16	50	72
17	V17	50	72
18	V18	44	67
19	V19	47	67
	Average	47	68

As presented in Table 1, the average pretest score obtained by students is 47, while the posttest score is 68. There is an increase in value between the pretest and posttest scores found in the vlog project class.

Table 2. Pre-test and post-test of picture series class

No.	Student	Pre-test	Post-test
1	P1	48	67
2	P2	44	61
3	P3	47	66
4	P4	46	63
5	P5	48	66
6	P6	46	64
7	P7	44	60
8	P8	46	65
9	P9	49	67
10	P10	48	66
11	P11	50	70
12	P12	50	69
13	P13	52	71
14	P14	52	71
15	P15	44	63

16	P16	45	63
17	P17	44	60
18	P18	48	68
	Average	47	66

In Table 2, the average pretest score obtained by students is 47, while the posttest score is 66. Similar to the vlog project class, the picture series class also experienced an increase between the pretest and posttest scores.

Table 3. Pre-test and post-test of conventional class

No.	Student	Pre-test	Post-test
1	C1	44	59
2	C2	46	61
3	C3	44	59
4	C4	48	63
5	C5	50	68
6	C6	50	66
7	C7	44	60
8	C8	50	66
9	C9	49	65
10	C10	48	62
11	C11	44	59
12	C12	45	61
13	C13	48	64
14	C14	50	68
15	C15	46	61
16	C16	48	65
17	C17	47	62
18	C18	46	61
19	C19	50	66

20	C20	45	60
	Average	47	63

In Table 3, the average pretest score obtained by students is 47, while the posttest score is 63. The learning process increased the performance of all classes.

Clustering Pre-test

The method section explains that the R programming language assists this research. The libraries and functions used to apply K-Means are the cluster library and the kmeans() function. Each class will be grouped into three groups: cluster 1 = group of students with low pre-test scores, cluster 2 = group of students with medium pre-test scores, and cluster 3 = group of students with high pre-test scores. The results of student clustering for pre-test scores have been presented in Table 4.

Table 4. Pre-test clustering results

Class	Cluster	Student	Average student grades
Vlog Project	Cluster 1	V1, V2, V3, V4, V5, V7, V11, V12, V14, V18	45
	Cluster 2	V6, V10, V13, V19	48
	Cluster 3	V8, V9, V15, V16, V17	50
Picture Series	Cluster 1	P2, P4, P6, P7, P8, P15, P16, P17	45
	Cluster 2	P10, P3, P5, P9, P11, P18	48
	Cluster 3	P12, P13, P14	51
Conventional	Cluster 1	C1, C2, C3, C7, C11, C12, C15, C18, C20	45

	Cluster 2	C4, C9, C10, C13, C16, C17	48
	Cluster 3	C5, C6, C8, C14, C19	50

Clustering Post-test

As with pretest scores, post-test scores were also grouped to determine any differences between groups formed based on the pretest and based on post-test scores. The results of student grouping for post-test scores have been presented in Table 5.

Table 5. Post-test clustering results

Class	Cluster	Student	Average student grades
Vlog Project	Cluster 1	V7	60
	Cluster 2	V1, V2, V3, V4, V5, V11, V14, V18, V19	66
	Cluster 3	V6, V8, V9, V10, V12, V13, V15, V16, V17	71
Picture Series	Cluster 1	P2, P4, P7, P15, P16, P17	62
	Cluster 2	P1, P3, P5, P6, P8, P9, P10, P18	66
	Cluster 3	P11, P12, P13, P14	70
Conventional	Cluster 1	C1, C2, C7, C11, C20	59
	Cluster 2	C2, C4, C10, C12, C15, C17, C18	62
	Cluster 3	C5, C6, C8, C9, C13, C14, C16, C19	66

Table 4 and Table 5 show a difference between the average pre-test and post-test scores. Table 5 shows that the number of students included in Cluster 1 is less than the other clusters in all classes. The highest average post-test score for cluster 3 is in the vlog project class. The vlog project class has an average post-test score of 71, the picture series score of 70, and the conventional class has a score of 66.

Cluster Verification

Each student is identified in cluster verification to see if there is a change after the learning process. The identification process is done by looking at changes in student clusters based on pre-test and post-test scores. The results of cluster verification can be seen in Table 6.

Table 6. Cluster verification

Class	Student	Cluster origin	Move to cluster
Vlog Project	V1	Cluster 1	Cluster 2
	V2	Cluster 1	Cluster 2
	V3	Cluster 1	Cluster 2
	V4	Cluster 1	Cluster 2
	V5	Cluster 1	Cluster 2
	V11	Cluster 1	Cluster 2
	V12	Cluster 1	Cluster 2
	V14	Cluster 1	Cluster 2
	V18	Cluster 1	Cluster 2
	V6	Cluster 2	Cluster 3
	V10	Cluster 2	Cluster 3
	V13	Cluster 2	Cluster 3
Picture Series	P6	Cluster 1	Cluster 2
	P8	Cluster 1	Cluster 2
	P11	Cluster 2	Cluster 3
Conventional	C3	Cluster 1	Cluster 2
	C12	Cluster 1	Cluster 2
	C15	Cluster 1	Cluster 2
	C18	Cluster 1	Cluster 2

C9	Cluster 2	Cluster 3
C13	Cluster 2	Cluster 3
C16	Cluster 2	Cluster 3

In Table 6, the class that experienced the most cluster changes was the vlog project class. Students who experience cluster movement make up 12 or approximately 63% of the students in the vlog project class. In the picture series class, students who experienced cluster changes were three students, and for conventional classes, there were as many as seven classes.

4 Conclusions

This study has successfully analyzed the best learning method to improve speaking skills by using the K-Means algorithm. The results of the cluster verification show that the vlog project is the class whose students have the most cluster changes compared to the other classes. That can be a reference to the fact that the vlog project is one of the effective learning techniques for improving speaking skills. This study only uses two assessment parameters to determine the best learning method. It is hoped that subsequent studies can use other parameters to strengthen the results of this study.

Acknowledgements

The author would like to thank the staff of MTs Negeri 13 Ciamis who have supported this research so that this research can be completed.

References

[1] A. E. Warliati, Z. Rafli, and D. Darmahusni, “Discussion and Think-Pair-Share Strategies on the Enhancement of EFL Students’ Speaking Skill: Does Critical Thinking Matter?,” *Journal of English Language Studies*, vol. 4, no. 2, p. 120, Sep. 2019, doi: 10.30870/jels.v4i2.6100.

[2] Riris Nurkholidah Rambe, Andini Syahfitri, Aini Humayroh, Nadila Alfina, Putri Azkia, and Tania Dwi Rianti, "Upaya Meningkatkan Keterampilan Berbicara Di Depan Umum," *Jurnal Pendidikan dan Sastra Inggris*, vol. 3, no. 2, pp. 11–24, Jun. 2023, doi: 10.55606/jupensi.v3i2.1966.

[3] H. Hotmaria, "Upaya Meningkatkan Keterampilan Berbicara Bahasa Inggris pada Materi Pengandaian Diikuti Perintah/Saran Menggunakan Strategi Pembelajaran Three Step Interview," *Journal of Education Action Research*, vol. 5, no. 1, Jan. 2021, doi: 10.23887/jear.v5i1.31558.

[4] I. Hikmawan, I. L. Damayanti, and S. Setyarini, "Integrating Traditional Games into EFL Speaking Class: A case of English for Young Learners," *Journal of English Language Studies*, vol. 8, no. 2, p. 249, Sep. 2023, doi: 10.30870/jels.v8i2.19432.

[5] N. H. Wulandari and A. Ashadi, "Vlog Project or Picture Series: Examining Effective Techniques in Teaching Speaking Skills," *Jurnal Pendidikan Progresif*, vol. 11, no. 2, pp. 275–289, 2021, doi: 10.23960/jpp.v11.i2.202111.

[6] Ana Sri Lestari and N. I. Sholichah, "Improving Speaking Ability By Using Picture Series," *Jurnal Penelitian Ilmiah INTAJ*, vol. 6, no. 1, pp. 22–41, Apr. 2022, doi: 10.35897/intaj.v6i1.728.

[7] Chairani Annisa, Reflinda Reflinda, Melyan Melani, and Veni Roza, "The Effect of Using Picture Series toward Students' Speaking Skill at the Second Grade in MTSS Nagari Binjai," *Populer: Jurnal Penelitian Mahasiswa*, vol. 1, no. 4, pp. 19–33, Dec. 2022, doi: 10.58192/populer.v1i4.248.

[8] Z. Zubaidi, R. P. Suharto, and E. L. Rahayu, "Improving Students' Speaking Skill through Students Vlog Project as PBL Output on Online Speaking Class," *Briliant:*

Jurnal Riset dan Konseptual, vol. 6, no. 4, p. 764, Nov. 2021, doi: 10.28926/briliant.v6i4.757.

[9] K. Febianti, S. Wahyuni, and I. Imadona, "Vlog Project In Enhancing Students' Speaking Achievement," *Esteem Journal of English Education Study Programme*, vol. 6, no. 2, pp. 353–360, Jul. 2023, doi: 10.31851/esteem.v6i2.12321.

[10] B. Wu and C. Zheng, "An Analysis of the Effectiveness of Machine Learning Theory in the Evaluation of Education and Teaching," *Wirel Commun Mob Comput*, vol. 2021, pp. 1–10, Oct. 2021, doi: 10.1155/2021/4456222.

[11] S. Anwar, T. Suprapti, G. Dwilestari, I. Ali, P. Studi Rekayasa Perangkat Lunak Jln Perjuangan No, and B. Kesambi Kota Cirebon, "Pengelompokkan Hasil Belajar Siswa Dengan Metode Clustering K-Means Program Studi Sistem Informasi Jln Perjuangan No 10B Kesambi Kota Cirebon 4)," *Jurnal Sistem Informasi dan Teknologi Informasi)*, vol. 4, no. 2, pp. 60–72, 2022.

[12] S. N. Br Sembiring, H. Winata, and S. Kusnasari, "Pengelompokan Prestasi Siswa Menggunakan Algoritma K-Means," *Jurnal Sistem Informasi Triguna Dharma (JURSI TGD)*, vol. 1, no. 1, p. 31, Jan. 2022, doi: 10.53513/jursi.v1i1.4784.

[13] V. Purwayoga and B. Susanto, "Pengelompokkan Daerah Berdasarkan Ketersediaan Masjid Muhammadiyah Dengan Algoritma K-Means," vol. 13, no. 1, 2021, doi: 10.24853/jurtek.13.1.75-80.

[14] V. Purwayoga, "Optimasi Jumlah Cluster pada Algoritme K-Means untuk Evaluasi Kinerja Dosen," *Jurnal Informatika Universitas Pamulang*, vol. 6, no. 1, p. 118, Mar. 2021, doi: 10.32493/informatika.v6i1.9522.

[15] H. Gunawan and V. Purwayoga, "Data Mining Menggunakan Algoritma K-Means Clustering Untuk Mengetahui Potensi Penyebaran Virus Corona Di Kota Cirebon,"

Jurnal Sisfokom (Sistem Informasi dan Komputer), vol. 11, no. 1, pp. 1–8, Jan. 2022, doi: 10.32736/sisfokom.v11i1.1316.

[16] P. Pangestu, S. Maarip, Y. Nur Addinsyah, and V. Purwayoga, "Clustering and Trend Analysis of Priority Commodities in the Archipelago Capital Region (IKN) using a Data Mining Approach," *International Journal of Applied Sciences and Smart Technologies*, vol. 6, no. 1, Jun. 2024, doi: 10.24071/ijasst.v6i1.7798.