PAPER NAME

Proceeding eLFA (ISBN 20072022) (2).pdf

WORD COUNT	CHARACTER COUNT
2371 Words	**12847 Characters**
PAGE COUNT	FILE SIZE
10 Pages	**1.6MB**
SUBMISSION DATE	REPORT DATE
Oct 30, 2023 3:21 PM GMT+7	**Oct 30, 2023 3:21 PM GMT+7**

5% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

- 3% Internet database
- Crossref database
- 3% Submitted Works database
- 0% Publications database
- Crossref Posted Content database

Excluded from Similarity Report

- Bibliographic material
- Cited material
- Manually excluded sources
- Quoted material
- Small Matches (Less then 10 words)
- Manually excluded text blocks

Unika SOEGIJAPRANATA
Talenta pro patria et humanitate

Proceeding
eLearning Forum Asia 2021

Augmenting the Virtual Environment: Technology | Innovation | Humanity

Semarang, 30 November - 1 December 2021

PeLS Philippine eLearning Society
SYMPLICITY
UNITED BOARD
eLearning Forum Asia

Editor:
Dr. Heny Hartono, SS, M.Pd
Cecilia Titiek Murniati, PhD

Website:
https://www.elfasia.org/2021/

Soegijapranata Catholic University

Proceeding
eLearning Forum Asia 2021

Augmenting the Virtual Environment: Technology – Innovation – Humanity

Editor

Dr. Heny Hartono, SS, M.Pd
Cecilia Titiek Murniati, PhD

Semarang, 30 November - 1 December 2021

Website
https://www.elfasia.org/2021/

Soegijapranata Catholic University

Proceeding
eLearning Forum Asia 2021

Augmenting the Virtual Environment: Technology – Innovation – Humanity

Editor:

Dr. Heny Hartono, SS, M.Pd
Cecilia Titiek Murniati, PhD

ISBN 978-623-5997-14-8 (PDF)

Cover Design	: Manggar
Layout	: Ernanto
Book size	: 21,5 x 30,35 cm
Font	: Goudy Old Style 12
Website	: https://www.elfasia.org/2021/
Presentation Guidelines	: ZOOM - https://www.elfasia.org/2021/?page_id=2773

Publisher:
Universitas Katolik Soegijapranata
Anggota APPTI No. 003.072.1.1.2019
Anggota IKAPI No 209/ALB/JTE/2021
Jl. Pawiyatan Luhur IV/1 Bendan Duwur Semarang 50234
Telpon (024)8441555 ext. 1409
Website: www.unika.ac.id
Email Penerbit: ebook@unika.ac.id

Unika SOEGIJAPRANATA
Talenta pro patria et humanitate

Proceeding
eLearning Forum Asia 2021

Augmenting the Virtual Environment: Technology | Innovation | Humanity

Semarang, 30 November - 1 December 2021

PeLS Philippine eLearning Society · SYMPLICITY · UNITED BOARD · eLearning Forum Asia

Assesment of e-Learning

Editor:
Dr. Heny Hartono, SS, M.Pd
Cecilia Titiek Murniati, PhD

Website:
https://www.elfasia.org/2021/

Soegijapranata Catholic University

Teachers' Language Competence Segmentation Using K-Mean Algorithm

[1]Rosita Herawati, [2]Heny Hartono and [3]Cecilia Titiek Murniati

[1]rosita@unika.ac.id, [2]heny@unika.ac.id, [3]c_murniati@unika.ac.id

[1, 2, 3]Soegijapranata Catholic University, Indonesia

Abstract: This paper presents a part of an educational research and development which has been being conducted for the last three years with a focus on bilingual program teachers' communicative competence assessment. This study involved six aspects of communicative competence by means of three di erent assessments, so called self-assessment, receptive productive assessment, and performance assessment. Despite the fact that the assessment has been digitalized, due to the pandemic time, only self-assessment could be done. The other two kinds of assessment could not be done because of the limited access to the schools during the pandemic.

This study highlighted the teachers' segmentation in 3 di erent classes. This segmentation is done using the K-means algorithm. This algorithm is vastly used for clustering in many data science applications. The Elbow as a common heuristic mathematical method is also used to optimize these algorithms. The optimum K value is modeled using a metric WCSS (Within Cluster Sum of Squares).

The teachers' language competence segmentation study is the practice of partitioning a teacher into groups that have similar language competence. It is a signi cant strategy to recognize teachers' language competence. The results of the study can be used to map teachers' communicative competence as well as giving recommendations for the most appropriate professional teacher training to improve their communicative competence.

Keywords: communicative competence assessment, K-means algorithm, Elbow, teachers' language competence

INTRODUCTION

The COVID-19 pandemic has changed many things in Indonesia. Activities that were supposed to be conducted face-to-face had to be postponed or even canceled, because at that time the Indonesian government had imposed large-scale social restrictions. Likewise, teaching and learning activities in schools must be temporarily suspended. This also has an impact on communicative competence assessment research activities that are focused on the competence of bilingual school teachers.

The bilingual teacher who is the subject of this research will be assessed with 3 assessment components. These components are Self-Re ection, Receptive-Productive, and Performance Assessment. Self-re ection assessment is the stage where the teacher is asked to assess himself. This assessment is done by lling in the rubric provided. While the other 2 assessments were carried out by assessors. Despite the fact that the assessment has been digitalized, due

to the pandemic time, only self-assessment could be done. While the others assessment cannot be carried out during the pandemic because of the limited access to the schools during the pandemic.

Although only self-assessment can be done during a pandemic, this research can still be done. Because each component of the assessment has accommodated the six aspects that emphasize teacher's competence. The six aspects are "linguistic competence, sociolinguistic competence, strategic competence, formulaic competence, discourse competence, and interactional competence"(Celce-Murcia, 2007, pp. 41-57). Although it is not perfect as if it was done with 3 components of assessment, with these six aspects, the nal impression can still be obtained.

The nal impression obtained from the results of the self-assessment will be segmented into 3 groups that have similar characteristics. This segmentation is done using the K-means algorithm. This algorithm is vastly used for clustering in many data science applications. The Elbow as a common heuristic mathematical method is also used to optimize these algorithms. The optimum K value is modeled using a metric WCSS (Within Cluster Sum of Squares). It is a signi cant strategy to recognize teachers' language competence. The results of the study can be used to map teachers' communicative competence as well as giving recommendations for the most appropriate professional teacher training to improve their communicative competence.

DISCUSSION

Vankayalapati, R et al. (2021, p.99-104) state that K-means clustering is a data mining algorithm that can be used to group data without knowing the relationship between data rstly. "K-means is an old cluster algorithm created by MacQueen in 1967 that is most commonly used. It was one of the easiest, unattended learning algorithms to solve the issue of the well-known cluster" (Shamir & Sharan, 2002, p.269-300).

The use of K-means in education and assessment is very common. Shahiri and Husain (2015, p.414-422) evaluated student gures in order to forecast student performance. The clustering of data can be used as k-means to assess the performance of students (Arora & Badal, Dr., 2013)

The basic idea of K-means algorithms is to de ne k centroids, one for each group. Then choose a point randomly as many as K numbers de ne, where this point will be the center(centroid) of each group (cluster). This is done randomly because later, the K-Means algorithm will change the position of each point until the optimal solution is reached. The next step is to create the dataset that is closest to the center point as part of the cluster. So that in total it will K clusters be formed. Based on the grouping in the second step, each data point will be joined in a cluster. Perform calculations, and place a new centroid center for each cluster. This step can also be referred to as centroid re nement.

The step will be continued by performing calculations according to the K-Means algorithm, i.e. nding the most suitable center point position for each cluster based on the closest distance calculation. Calculation of the distance of each data point to the center of the cluster can use the Euclidean distance method. From the dataset that we have, we take the nearest centroid point, so that the dataset becomes part of the cluster. If there is still data that changes groups (move cluster), the previous step must be retake.

Otherwise, the cluster is formed already well.

The last step of the K-Means algorithm is to check the de ned center point. Select the nearest center point, and enter into the cluster. If there is still a cluster shift, repeat the previous step. The K-Means algorithm will continue to nd the center point, until the division of the dataset has been optimized and the position of the midpoint has not changed anymore.

A. K-means Optimation

Upon the discussion above, it can be analyzed that one of the crucial factors is determining the number of clusters (K value). Due to results, grouping will result in a di erent analysis for the number of clusters. If K is too few, then cluster division becomes fast, but it is possible there is hidden information that is not revealed. If K is too big, then the number of clusters is too many. It might be too di cult to choose a strategy that is appropriate for each cluster.

To overcome this, an optimization function can be added which will select the number of the start of the cluster correctly. (Dabbura Imad, 2018) The elbow method will help to choose the right K value by using metric WCSS (Within Cluster Sum of Squares). As an example, below is a calculation for two clusters:

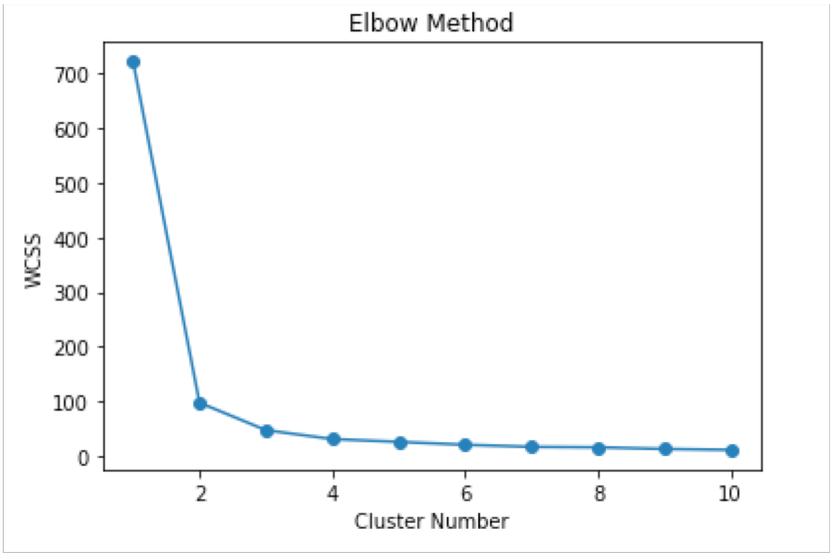
$$\text{WCSS} = \sum_{P_i \text{ in Cluster 1}} \text{distance}(P_i, C_1)^2 + \sum_{P_i \text{ in Cluster 2}} \text{distance}(P_i, C_2)^2 + \sum_{P_i \text{ in Cluster 3}} \text{distance}(P_i, C_3)^2$$

In the metric above, it can be observed that WCSS is the dependent variable. The Sigma symbol represents the sum of the squares of the distances of each point Pi in cluster 1. The sum of squares is the sum of the squares of each distance. Next, the results of cluster 1 are added with the result of the square of the distance of each data point to the center point of cluster two, and and so on according to the number of clusters we want.

To nd out the best number of clusters for the case study being tested is to look at the comparison of WCSS for 2 clusters, 3 clusters, 4 and carry on. What we choose is when the change in WCSS value is very signi cant, like elbows. Therefore this selection method is called the elbow method. And the resulting Elbow method upon the self assessment that participants has been done is shown below:

Figure 1:
Optiomization of K_means with the elbow method to determine the right number of clusters
has been done is shown below:

Figure 1:
Optimization of K-Means with the elbow method to determine the right number of clusters

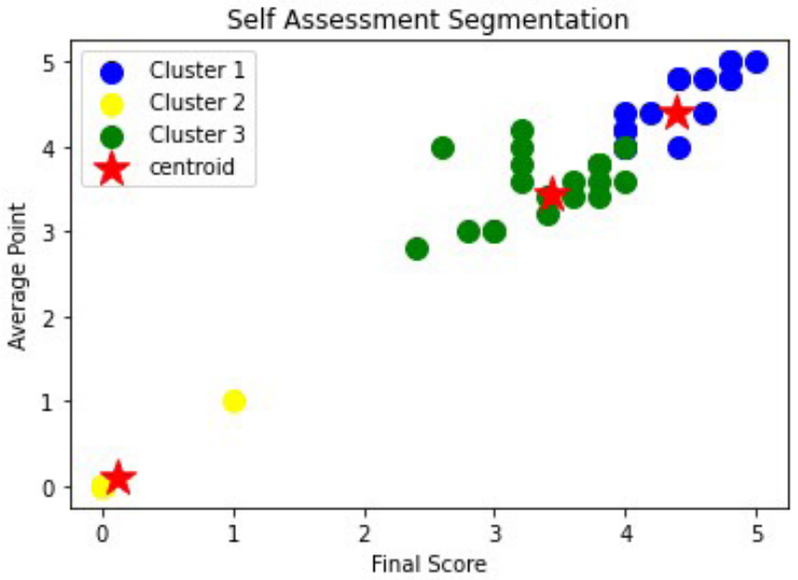


The estimated value of K is determinted by looking at the change in the value on the graph. The WCSS value indicated that after the number 3, the changes are no longer signi cant.

B. Result

The gure 1 shows the optimization K-means with the elbow method. The elbow curve shows that we need 3 centroids to grouping the data. Because And by using the optimal cluster value above, a model is developed to perform the classi cation. The gure 2 below shows the classi cation of self assessment with six aspects of aspects of a teacher's competence.

Figure 2
Visualization of K-Means with K=3



In this study, there are six diagrams of self-assessment results with 6 aspects of a teacher's competence. However, the six diagrams have the same pattern. For this reason, in this paper only one image is presented as a representative of the other diagrams.

This diagram is a visualization of the nal score and average score obtained by participants who have lled out the self-assessment matrix. The nal score will be mapped with the band descriptor as the nal impression of overall assessment. The table showing the band descriptors can be seen below. This table is taken from the assesment book which is used as a guide in this study.

Table 1:
The table of band descriptors

Final Score	Description
1	Intermittent Communicator : Communication occurs only sporadically.
2	Limited Communicator : Receptive/productive skills do not allow continuous communication.
3	Moderate Communicator : Gets by without serious breakdowns. However, misunderstandings and errors cause difficulties.
4	Competent Communicator : Copes well but has occasional misunderstandings or makes occasional noticeable errors.
5	Good Communicator : Copes well and performs competently.

The diagram above shows that participants who are bilingual teachers already have su cient competence.

This can be seen in 2 large clusters which show a nal score between 3 and 5. If it is mapped on the band descriptor, then most of the participants, approximately 50%, are included in the Moderate Communicator. While the remaining 34% have got the nal impression as a Competent Communicator.

CONCLUSION

K means as a clustering algorithm that depends on the number of clusters, relies on other methods such as Elbow to determine the optimal number of clusters. The number of clusters depends on the manual identi cation of the elbow point on the visualization graph. The self assessment data is segmentation into 3 clusters. The segmentation shows 48.7% of Medium Communicators, the other 46.2% are Competent Communicators, and around 5.1% are included in Limited Communicators.

To support the professional development of teachers, assessment of communicative competence is needed. This assessment will show the level of teacher competence. By knowing the level of teacher competence, and segmented them into similar groups, teacher's training can be tailored to the needs of each cluster. training can be designed to be targeted and more e ective.

ACKNOWLEDGEMENT

This research is a part of multi-year research supported by the Indonesian Ministry of Higher Education (Dikti). The writers would like to extend their gratitude to DIKTI which has nancially support this research.

REFERENCES

Arora, Rakesh & Badal, Dr. (2013). Evaluating Student's Performance Using k-Means Clustering. www.ijcst.com. 4.

Dabbura Imad. (2018). K-means Clustering - Algorithm, Applications, Evaluation Methods, and Drawbacks. Retrieved from https://imaddabbura.github.io/post/kmeans-clustering/#evaluat ion

M. Celce-Murcia, "Rethinking the communicative competence in language teaching", in E. A. Soler, & M. P. Jorda, Intercultural Language Use and Language Learning, Dordrecht: Springer, 2007, pp. 41-57.

Shahiri, A.M., Husain, W. (2015). A review on predicting student's performance using data mining techniques. Procedia Computer Science, 72: 414-422. https://doi.org/10.1016/j.procs.2015.12.157

Shamir R, Sharan R: Algorithmic approaches to clustering gene expression data. In Current Topics in Computational Molecular Biology. MIT Press; 2002:269–300.

Vankayalapati, R., Ghutugade, K.B., Vannapuram, R., Prasanna, B.P.S. (2021). K-means algorithm for clustering of learners performance levels using machine learning techniques. Revue d'Intelligence Arti cielle, Vol. 35, No. 1, pp. 99-104. https://doi.org/10.18280/ria.350112

The 16th eLearning Forum Asia 2021 was a collaborative event organized by Soegijapranata Catholic University, E-learning Forum Asia, and the United Board. The conference is usually held in different institutions within Asia regions. In 2021, eLearning Forum Asia took place in Soegijapranata Catholic University, Semarang, Indonesia. The conference theme, Augmenting the Virtual Environment: Technology – Innovation – Humanity, was to invite researchers, technologists, educators, and students to share virtual experiences, showcase teaching innovations, share insights and ideas on balancing teacher and students' needs and teaching goals.

ISBN 978-623-5997-14-8 (PDF)

5% Overall Similarity

Top sources found in the following databases:

- 3% Internet database
- Crossref database
- 3% Submitted Works database
- 0% Publications database
- Crossref Posted Content database

TOP SOURCES

The sources with the highest number of matches within the submission. Overlapping sources will not be displayed.

1	**old.unika.ac.id** Internet	2%
2	**Universitas Katolik Indonesia Atma Jaya on 2022-05-20** Submitted works	2%
3	**Institute of Technology Blanchardstown on 2013-12-15** Submitted works	1%
4	**National University of Singapore on 2016-04-08** Submitted works	<1%

Excluded from Similarity Report

- Bibliographic material
- Cited material
- Manually excluded sources
- Quoted material
- Small Matches (Less then 10 words)
- Manually excluded text blocks

EXCLUDED SOURCES

repository.unika.ac.id 87%

Internet

EXCLUDED TEXT BLOCKS

eLearning Forum Asia 2021Augmenting the Virtual Environment:Technology - Inno...

old.unika.ac.id

eLearning Forum Asia 2021Augmenting the Virtual Environment:Technology - Inno...

old.unika.ac.id

All rights reserved. Reproduction or transfer of part or all of the contents in this bo...

repository.unimus.ac.id

Universitas Katolik SoegijapranataAnggota APPTI No. 003.072.1.1.2019Anggota I...

id.123dok.com