Journal of Software Engineering, Information and Communication Technology (SEICT)

Journal homepage: https://ejournal.upi.edu/index.php/SEICT

CLUSTER ANALYSIS OF EMOTIONS IN QURANIC TRANSLATIONS USING K-MEANS CLUSTERING

Muhamad Faisal Fiqri[1], Raditya Muhammad[2], Mochamad Iqbal Ardimansyah[3]

[1,2,3]Software Engineering, Universitas Pendidikan Indonesia, Indonesia

Correspondence: E-mail: mfaisal@upi.edu

ABSTRACT

Al-Qur'an as the word of Allah is a comprehensive source of knowledge, covering spiritual, moral, social, and psychological aspects, including instructions on the recognition of emotions that have a significant impact on a person's emotional intelligence. This research aims to identify and categorize verses in Indonesian translation of the Quran that contain basic emotions such as anger, disgust, fear, happiness, sadness, and surprise. The process involves data preprocessing, verse search using Vector Space Model, and application of K-Means Clustering algorithm. As a result, the verses can be grouped into four main clusters. The characteristics of the clusters formed include, cluster 0 shows the grouping of verses containing the word "happy", clusters 1 and 2 respectively show the word "fear", and cluster 3 shows the word "sad". The cluster evaluation results obtained using Silhouette Score is 0.442 and Calinski-Harabasz Index is 251.653, which indicates that there is a sign of cluster but there is still some overlap between clusters. In conclusion, this clustering makes an important contribution to the understanding of Quranic interpretation and opens up opportunities for further development in academic studies and religious learning

ARTICLE INFO

Article History:
Submitted/Received
9 October 2024
First Revised 18 October 2024
Accepted 22 November 2024
First Available online
11 December 2024
Publication Date
11 December 2024

Keyword:
K-Means Clustering,
Basic Emotion,
Al-Qur'an Translation,
Text Mining.

1. INTRODUCTION

The Qur'an as the word of God is an invaluable source of knowledge, covering various aspects of life such as spiritual, moral, social, and psychological (I. Idaman dan S. Hidayat, 2011) (A. Mustofa,2018) (S. Suparlan,2008). Its verses not only provide guidance, but also inspiration and motivation in living everyday life (A. Nurrohim dan I. N. Sidik, 2020). The wealth of information in the Qur'an makes it a source that continues to be explored, including in understanding emotions and psychological well-being.

The ability to regulate emotions is a critical life skill with a positive impact on adult life (R. E. Martin dan K. N. Ochsner,2016). Emotional intelligence has a significant negative influence on quarter life crisis in early adulthood, with a contribution of 83.7%. Of the 400 samples, 55.7% had high emotional intelligence and 56.2% experienced low quarter life crisis. Emotional intelligence also plays an important role in reducing feelings of helplessness and anxiety in early adulthood (I. L. Anggraeni dan Y. A. Rozali, 2024). Efforts to improve emotional intelligence can be done through reading literature

Technology plays an important role in Qur'anic interpretation in Indonesia, especially in facilitating access to and understanding of complex religious literature. Technology can be used as a tool to collect similar-themed verses and disseminate interpretations audiovisually (D. I. A. Putra dan M. Hidayaturrahman, 2020). Thematic interpretation methods supported by technology have proven to be more practical and faster than traditional methods. Clustering translations of Qur'anic verses with algorithms such as k-means allows identification of the best clusters, although further development is needed to focus on specific topics (Y. E. Saida, 2007). In addition, the application of similarity calculation methods is important to measure the similarity between Qur'ānic documents, which supports the effectiveness of search systems based on specific themes.

Based on this background, this study aims to explore the clustering of Qur'anic verse translations based on basic emotion topics using the k-means algorithm. Visualization of cluster results will be done with wordcloud for each cluster. This research is expected to provide an overview of the basic emotional topics discussed in the translation of the Qur'anic verses

2. METHOD

The research design procedure from start to finish is shown in Figure 2

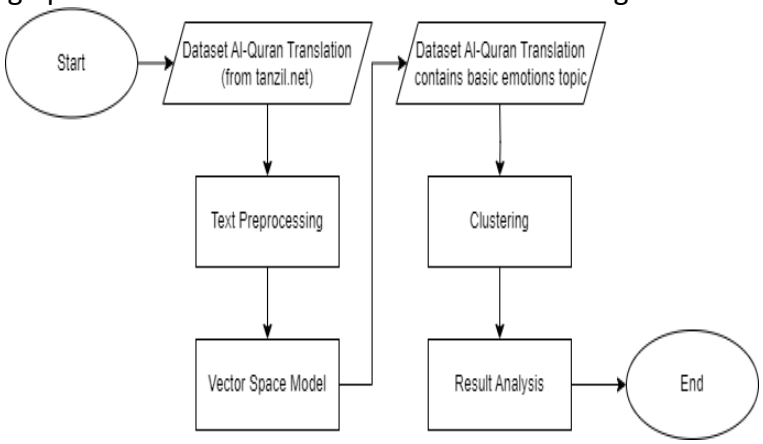


Figure 1. Research Procedure

2.1 Dataset

The dataset used in the research is a dataset of the translation of the Qur'an in Indonesian which amounts to 6236 verses. The dataset was obtained through the website

DOI: https://doi.org/10.17509/seict.v5i2.75942
p-ISSN 2774-1656 | e- ISSN 2774-1699

https://tanzil.net/trans/ compiled by the "Indonesian Ministry of Religious Affairs" uploaded on June 04, 2010. The downloaded dataset has a .txt format. It contains the translation of the Qur'an along with the letter and verse numbers separated by a "|" sign. The dataset will then be read using the pandas library provided by Python

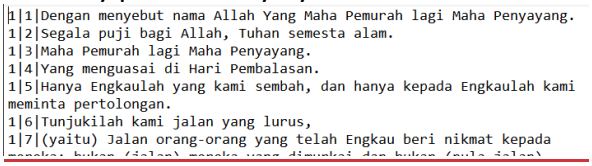


Figure 2. Dataset Quran Translation

2.2 Text Preprocessing

The stages of preprocessing are shown in the diagram below

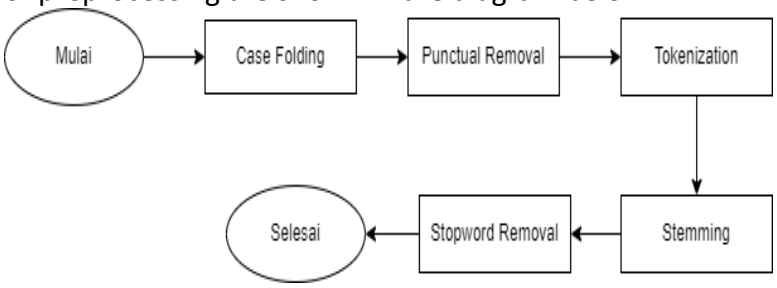


Figure 3. Prepocecssing Stages

The first step in data preprocessing after reading the dataset of all Al-Qur'an translations is case folding, which is the process of changing all characters in the text to lowercase letters to ensure uniformity and reduce variations caused by differences in capitalization. Next, punctuation removal is performed, which aims to remove or replace punctuation marks and symbols contained in the text, so that the analysis can be more focused on meaningful content. The next stage is tokenization, which is the separation of text into individual words or tokens based on spaces or punctuation, to prepare the data for further analysis. Next is stemming, which is the process of removing affixes or suffixes from a word so that it becomes the basic form. This process is very important in an effort to understand the basic meaning of words despite variations in their form. Finally, stopwords are removed, which is the process of removing common words or conjunctions that often appear in the text but tend not to contribute significantly to the understanding of the text content, so that the analyzed content becomes more focused.

2.3 Vector Space Model

The stages in forming a Vector Space Model to search for verse translations based on the keywords entered are shown in the diagram below.

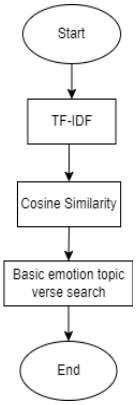


Figure 4. Vector Space Model Stages

DOI: https://doi.org/10.17509/seict.v5i2.75942
p-ISSN 2774-1656 | e- ISSN 2774-1699

The first stage is TF-IDF (Term Frequency-Inverse Document Frequency), which converts text data into a numerical representation by considering the frequency of occurrence of words in the document and their relevance across the document set. Next, Cosine Similarity is used to calculate the similarity value between two text vectors by measuring the cosine angle between them, which helps in identifying how similar the two texts are. The final stage involves searching for verses with basic emotion topics such as “anger,” “joy,” “sadness,” “fear,” “surprise,” and “disgust,” where a search is performed based on these keywords to find verses that correspond to those emotions.

2.4 Dataset of Qur'anic translations based on basic emotion topics

Form a new dataset based on translated Qur'anic verses containing the basic emotion keywords “senang sedih marah takut terkejut jijik”.

2.5 Clustering

Clustering stages are shown through the diagram below

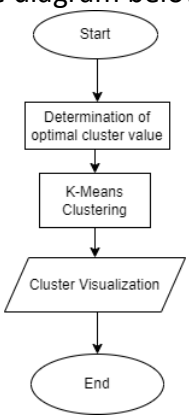


Figure 5. Clustering Stages

The first stage in this process is the search for optimal clusters, where the value of clusters (k) to be formed is determined using elbow method analysis. This method helps in identifying the right number of clusters by looking at the point where adding clusters no longer provides a significant reduction in variance. Once the optimal number of clusters is determined, K-means Clustering is performed using the k-means algorithm to form clusters based on the predetermined (k) value. The final step is cluster visualization, where the clustering results are visualized using PCA (Principal Component Analysis) plots to display the cluster distribution in two-dimensional space, as well as wordcloud visualization for each cluster to display the most dominant words in each cluster. This process allows for a more in-depth and intuitive analysis of the clustered data.

2.6 Result Analysis

In this study, the results will be analyzed using clustering metrics. The metrics used are Sillhouette Score and Calinski Harabasz Index. The value of the Silhouette Score can range from -1 to 1, the sillhoeutte score value can be interpreted as follows (L. Kaufman dan P. Rousseeuw, 1990):

Table 1. Sillhouette Score Interpretation

Score	Sillhouette Score Mean Interpretation
0.71 - 1.00	Strong structure found
0.51 - 0.70	Found a reasonable structure
0.26 - 0.50	Weak structure found
<0.25	No substantial structure found

DOI: https://doi.org/10.17509/seict.v5i2.75942
p-ISSN 2774-1656 | e- ISSN 2774-1699

3. RESULT AND DISCUSSION

3.1 Dataset

To start processing the Qur'an translation data, a dataset in .txt format downloaded from https://tanzil.net/trans/ will be used. This dataset will be read using the pandas library, which allows to convert text data into DataFrame format so that it is easier to analyze and process further. Pandas provides various functions that are very useful for manipulating, transforming, and exploring data, which will support the next steps in this research.

3.2 Preprocessing Data

a) Case Folding

Table 2. Case Folding Result

Input	*Output*
"Janganlah kamu bersikap lemah, dan janganlah (pula) kamu bersedih hati, padahal kamulah orang-orang yang paling tinggi (derajatnya), jika kamu orang-orang yang beriman."	"janganlah kamu bersikap lemah, dan janganlah (pula) kamu bersedih hati, padahal kamulah orang-orang yang paling tinggi (derajatnya), jika kamu orang-orang yang beriman."

b) Punctual Removal

Table 3. Punctual Removal Result

Input	*Output*
"janganlah kamu bersikap lemah, dan janganlah (pula) kamu bersedih hati, padahal kamulah orang-orang yang paling tinggi (derajatnya), jika kamu orang-orang yang beriman."	janganlah kamu bersikap lemah dan janganlah pula kamu bersedih hati padahal kamulah orang orang yang paling tinggi derajatnya jika kamu orang -orang yang beriman

c) Tokenization

Table 4. Tokenization Result

Input	*Output*
janganlah kamu bersikap lemah dan janganlah pula kamu bersedih hati padahal kamulah orang orang yang paling tinggi derajatnya jika kamu orang-orang yang beriman	['“janganlah'”, '“kamu'”, '“bersikap'”, '“lemah'”, '“dan'”, '“janganlah'”, '“pula'”, '“kamu'”, '“bersedih'”, '“hati'”, '“padahal'”, '“kamulah'”, '“orang'”, '“orang'”, '“yang'”, '“paling'”, '“tinggi'”, '“derajatnya'”, '“jika'”, '“kamu'”, '“orang”,”-orang'”, '“yang'”, '“beriman'”]

d) Stemming

Table 5. Stemming Result

Input	*Output*
['“janganlah'”, '“kamu'”, '“bersikap'”, '“lemah'”, '“dan'”, '“janganlah'”, '“pula'”, '“kamu'”, '“bersedih'”, '“hati'”, '“padahal'”, '“kamulah'”, '“orang'”, '“orang'”, '“yang'”, '“paling'”, '“tinggi'”, '“derajatnya'”, '“jika'”, '“kamu'”, '“orang’”, ‘“orang'”, '“yang'”, '“beriman'”]	['“jangan'”, '“kamu'”, '“sikap'”, '“lemah'”, '“dan'”, '“jangan'”, '“pula'”, '“kamu'”, '“sedih'”, '“hati'”, '“padahal'”, '“kamu'”, '“orang'”, '“orang'”, '“yang'”, '“paling'”, '“tinggi'”, '“derajat'”, '“jika'”, '“kamu'”, '“orang’”, ‘“orang'”, '“yang'”, '“iman'”]

DOI: https://doi.org/10.17509/seict.v5i2.75942
p-ISSN 2774-1656 | e- ISSN 2774-1699

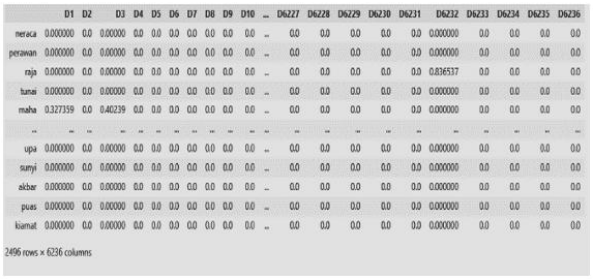
1. Stopword Removal

Table 6. Stopword Removal Result

Input	*Output*
['“jangan'”, '“kamu'”, '“sikap'”, '“lemah'”, '“dan'”, '“jangan'”, '“pula'”, '“kamu'”, '“sedih'”, '“hati'”, '“padahal'”, '“kamu'”, '“orang'”, '“orang'”, '“yang'”, '“paling'”, '“tinggi'”, '“derajat'”, '“jika'”, '“kamu'”, '“orang’”, ‘“orang'”, '“yang'”, '“iman'”]	['“sikap'”, '“lemah'”, '“sedih'”, '“hati'”, '“orang'”, '“derajat'”, '“iman'”]

2. Vector Space Model

a. TF-IDF



	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	...	D6227	D6228	D6229	D6230	D6231	D6232	D6233	D6234	D6235	D6236
neraca	0.000000	0.0	0.00000	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.0
perawan	0.000000	0.0	0.00000	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.0
raja	0.000000	0.0	0.00000	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.836517	0.0	0.0	0.0	0.0
tunai	0.000000	0.0	0.00000	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.0
maha	0.327359	0.0	0.40239	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
upa	0.000000	0.0	0.00000	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.0
sunyi	0.000000	0.0	0.00000	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.0
akbar	0.000000	0.0	0.00000	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.0
puas	0.000000	0.0	0.00000	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.0
kiamat	0.000000	0.0	0.00000	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.0

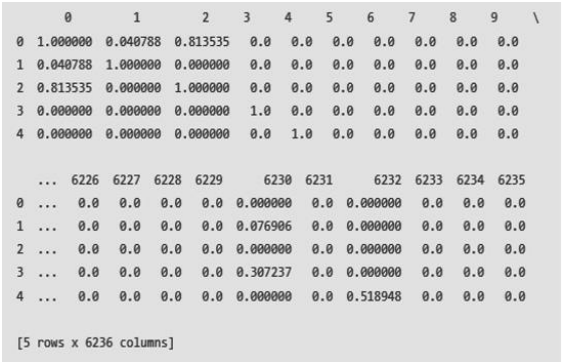
2496 rows × 6236 columns

Figure 6. Tf-Idf Result

Based On Figure 7, Showing Some Of The Tf-Idf Calculation Results Displayed Above Shows A Matrix Consisting Of 2496 Rows And 6236 Columns, Where Each Row Represents A Particular Word Or Term, And Each Column Represents A Document Or Paragraph In The Dataset. The Tf-Idf Value Calculated For Each Word In This Document Indicates How Important The Word Is In The Context Of That Particular Document.

b. Cosine Similarity

After obtaining the TF-IDF vector representation of each paragraph, the next step is to calculate the similarity between documents using Cosine Similarity. Cosine Similarity measures the similarity between two vectors by calculating the cosine of the angle between them. The result is a similarity matrix that describes how similar each document is to each other. This matrix is displayed in tabular form to provide a visual representation of the similarity between documents.



	0	1	2	3	4	5	6	7	8	9	\
0	1.000000	0.040788	0.813535	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
1	0.040788	1.000000	0.000000	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
2	0.813535	0.000000	1.000000	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
3	0.000000	0.000000	0.000000	1.0	0.0	0.0	0.0	0.0	0.0	0.0	
4	0.000000	0.000000	0.000000	0.0	1.0	0.0	0.0	0.0	0.0	0.0	

	...	6226	6227	6228	6229	6230	6231	6232	6233	6234	6235
0	...	0.0	0.0	0.0	0.0	0.000000	0.0	0.000000	0.0	0.0	0.0
1	...	0.0	0.0	0.0	0.0	0.076906	0.0	0.000000	0.0	0.0	0.0
2	...	0.0	0.0	0.0	0.0	0.000000	0.0	0.000000	0.0	0.0	0.0
3	...	0.0	0.0	0.0	0.0	0.307237	0.0	0.000000	0.0	0.0	0.0
4	...	0.0	0.0	0.0	0.0	0.000000	0.0	0.518948	0.0	0.0	0.0

[5 rows x 6236 columns]

Figure 7. Cosine Similarity Result

c. Basic Emotion Topic Verse Search

To find the most similar verse to the input text based on basic emotion topics, a function is created that receives input text from the user, preprocesses the text, and then converts it into a TF-IDF vector. This vector is then compared with the TF-IDF vector of all documents using Cosine Similarity. The document index with the highest similarity value is identified as the most similar document. The text used as keywords are “marah terkejut takut senang sedih jijik”.

DOI: https://doi.org/10.17509/seict.v5i2.75942
p-ISSN 2774-1656 | e- ISSN 2774-1699

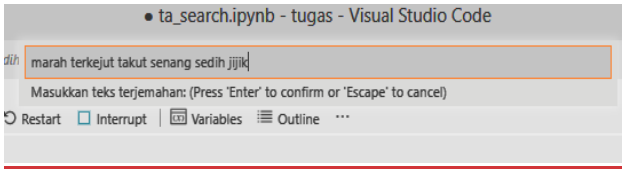


Figure 8. Basic Emotion Keyword Input For Qur'an Search

3. Dataset of Qur'anic translations based on basic emotion topics

After searching using the Vector Space Model, a new dataset is formed in the form of translations of Al-Qur'an verses containing basic emotion words according to the keywords that have been used. A total of 275 verses containing words on the topic of basic emotions were obtained.

	ID Dokumen	Surah	Ayat	Translation	Processed	Cosine Similarity
0	1498	11	26	agar kamu tidak menyembah selain Allah. Sesung...	sembah allah takut timpa azab sedih	0.310342
1	2433	20	86	Kemudian Musa kembali kepada kaumnya dengan ma...	musa kaum marah sedih hati hai tuhan janji asa...	0.286551
2	1292	9	58	Dan di antara mereka ada orang yang mencelamu ...	orang cela distribusi zakat sebahagian senang ...	0.265154
3	5544	74	50	seakan-akan mereka itu keledai liar yang lari ...	keledai liar lari kejut	0.256949
4	4308	42	37	Dan (bagi) orang-orang yang menjauhi dosa-dosa...	orang dosa keji marah maaf	0.252905
...	...	...	...	...	...	...
270	25	2	19	atau seperti (orang-orang yang ditimpa) hujan ...	orang timpa hujan lebat langit gelap gulita gu...	0.046218
271	3569	33	37	Dan (ingatlah), ketika kamu berkata kepada ora...	orang allah limpah nikmat tahan isterimu takwa...	0.044665
272	241	2	235	Dan tidak ada dosa bagi kamu meminang wanita-w...	dosa pinang wanita sindir sembunyi awin hati a...	0.038875
273	517	4	25	Dan barangsiapa diantara kamu (orang merdeka) ...	barangsiapa orang merdeka belanja awin wanita ...	0.035433
274	671	5	3	Diharamkan bagimu (memakan) bangkai, darah, da...	haram makan bangkai darah daging babi hewan se...	0.030971

275 rows × 6 columns

Figure 9. Basic Emotion Qur'an Translation Dataset

4. Clustering

a. Determination of Optimal cluster

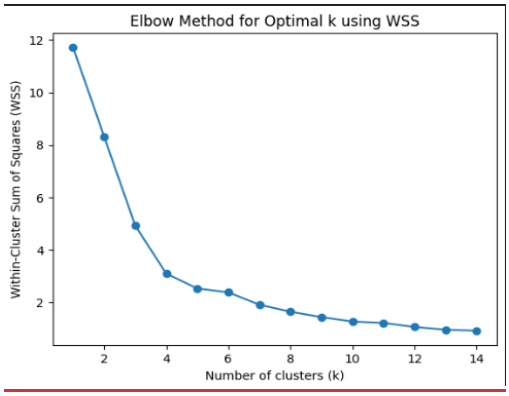


Figure 10. Elbow Method Plot

Based On The Elbow Plot In Figure 11, The Elbow Point Indicates The Optimal Number Of Clusters Because After This Point, The Decrease In Wss Becomes Less Significant Even Though The Number Of Clusters Increases. This Point Gives An Indication That Further Addition Of Clusters Does Not Give Any Significant Advantage In Terms Of Reducing The Variation In The Clusters, And Therefore, Is Considered As The Optimal Number Of Clusters For Data Clustering. Based On The Resulting Graph, The Elbow Point Is At A Value Of K=4 So That Value Will Be Used In Forming Clusters Using The K-Means Algorithm.

b. K-Means Clustering

After determining that the optimal number of clusters is 4 based on elbow analysis, the next step is to cluster the data using the K-means algorithm. By using the scikit-learn library, the implementation of this algorithm can be done easily. First, the K-means model is initialized by setting the number of clusters to 4. Then, the model will be trained on the dataset. The clustering result will be a cluster label for each data point, indicating which group the data belongs to.

DOI: https://doi.org/10.17509/seict.v5i2.75942
p-ISSN 2774-1656 | e- ISSN 2774-1699

c. Cluster Visualization

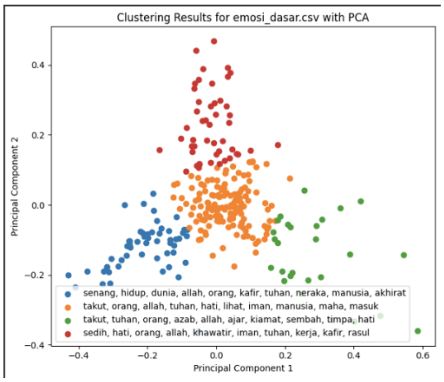


Figure 11. Cluster Visualization Using Pca

Through the visualization in Figure 12, there are 4 clusters formed according to the cluster value that has been determined based on the results of the elbow method analysis. In each cluster, the top 10 words in the Qur'anic translation are sorted from left to right. Each cluster contains words with basic emotion types such as happy, fear, and sadness as words that often appear in each cluster. For example, cluster 0 shown with blue data indicates that in this cluster, it contains verses that mention happy emotions. Cluster 1 and cluster 2 both show the word fear as a word that often appears in the cluster but with a little difference, one of which is that in cluster 2 which is green there are words of doom, doomsday which are not found in the top 10 words in cluster 1 which is orange. Cluster 3 in red shows the word sad as a word that often appears in the Qur'anic translation in that cluster.

Figure 12. Wordcloud Cluster 0

In cluster 0 shown in Figure 13, the word “happy” has the highest frequency with 51 occurrences, followed by the word “life” with 15 occurrences, and “world” with 14 occurrences. Other frequently occurring words include “allah” (12 occurrences), “people” (11 occurrences), “kafir” (10 occurrences), “god” (8 occurrences), “hell” (7 occurrences), “human” (7 occurrences), and “afterlife” (7 occurrences).

Figure 13. Wordcloud Cluster 1

In cluster 1 shown in Figure 14, the word “fear” appears most frequently with a frequency of 115 occurrences, followed by “people” with 97 occurrences, and “god” with 71 occurrences. Other words that also appear frequently include “god” (31 occurrences), “heart”

DOI: https://doi.org/10.17509/seict.v5i2.75942
p-ISSN 2774-1656 | e- ISSN 2774-1699

(28 occurrences), “faith” (21 occurrences), “see” (19 occurrences), “human” (17 occurrences), “happy” (15 occurrences), and “enter” (15 occurrences).

Figure 14. Wordcloud Cluster 2

In cluster 2 shown in Figure 15, the word “fear” appears most frequently with a frequency of 37 occurrences, followed by “god” with 27 occurrences, and “allah” with 14 occurrences. Other frequently occurring words include “doom” (17 occurrences), “apocalypse” (8 occurrences), “override” (6 occurrences), and “teach” (5 occurrences)

Figure 15. Wordcloud Clutser 3

In Cluster 3 Shown In Figure 16, The Word “Sad” Appears Most Frequently With A Frequency Of 35 Occurrences, Followed By “Heart” With 33 Occurrences, And “Person” With 22 Occurrences. Other Words That Also Appear Frequently Include “Allah” (21 Occurrences), “Worry” (16 Occurrences), “God” (9 Occurrences), “Faith” (8 Occurrences), And “Apostle” (8 Occurrences).

5. Result Analysis

a. Sillhouette Score

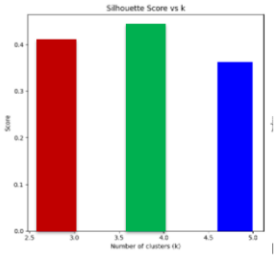


Figure 16. Sillhouette Score For K=3, K=4, And K=5

The Result Obtained Is That The Value Of K = 4 Has A Higher Silhouette Score (0.442) Compared To K = 3 (0.408) And K = 5 (0.362). Referring To The Interpretation Made By (Kaufman & Rousseeuw, 1990) The Silhouette Score Value Of 0.442 Indicates That There Is An Indication Of Clustering In The Data, But The Strength Of The Clustering Is Not Very Strong. In More Detail, This Value Indicates That Although Some Data Are Quite Close To The Centroid Of Their Cluster, There Are Still A Number Of Data That May Be On The Border Between Different Clusters, Or Even Outside The Supposed Cluster. In Other Words, The Clustering May Not Be Optimal, And There Is A Possibility Of Overlap Between Clusters. Overall, These Results Show That Although There Are Some Structures In The Data That Support Clustering, The Clustering Results Are Not Optimal And Can Be Further Improved.

DOI: https://doi.org/10.17509/seict.v5i2.75942
p-ISSN 2774-1656 | e- ISSN 2774-1699

b. Calinski Harabasz Index

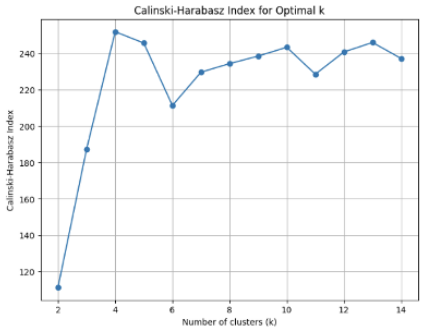


Figure 17. Ch Index From K=1 To K=14

Based On The Graph In Figure 18 Showing The Calinski Harabasz Index For Various Number Of Clusters (K), It Can Be Observed That The Highest Index Value Is Achieved When K=4, With A Value Of 251,653. This Indicates That Dividing The Data Into Four Clusters Results In Optimal Clustering Quality, According To The Calinski Harabasz Criterion, Which Measures How Well The Clusters Are Separated From Each Other And How Closely The Cluster Members Are Grouped.

When The Number Of Clusters Is Increased To K=6, The Index Value Decreases To 211.192, Which Is The Lowest Value In The Range Of Clusters Evaluated. This Decrease Indicates That Increasing The Number Of Clusters Beyond Four Not Only Fails To Improve The Quality Of Clusterization, But Also Potentially Worsens The Separation Between Clusters.

After K=4, Increasing The Number Of Clusters Does Not Result In A Significant Improvement In Clustering Quality, As Indicated By The Trend Of The Calinski Harabasz Index Which Tends To Be Stable With Little Variation. This Shows That Although There Are Small Fluctuations In The Index Values For Higher (K), Selecting A Number Of Clusters Greater Than Four Does Not Provide Any Significant Advantage In Clustering Quality. Thus, In This Context, Choosing K=4 As The Optimal Number Of Clusters Is The Right Decision Based On The Performance Measured By The Calinski Harabasz Index.

5. CONCLUSION

Based on the results of research on the clustering of translations of Qur'anic verses on the topic of basic emotions, it is concluded that through text mining approach using k-means cluster algorithm in grouping translations based on certain topics is successfully done. The findings obtained are as follows:

1. From the results of the implementation of k-means clustering by determining the number of cluster values using the elbow method and visualization using PCA and Wordcloud, an optimal number of four clusters is obtained which highlights three of the six types of basic emotions such as happy, sad, and fear emotions. Cluster 0 contains the translation of Qur'anic verses that contain words with the type of happy emotion as the most words, clusters 1 and 2 contain the word fear as the most words and cluster 3 contains the word sad as the most words.
2. The evaluation results of clustering Indonesian Qur'anic translations with basic emotion topics using k-means clustering show that the optimal number of clusters is four clusters (k=4). Evaluation analysis shows that the cluster with k=4 value produces the highest value of 0.442 for Silhouette Score and 251.653 for Calinski-Harabasz Index compared to other cluster numbers. Silhouette Score values below 0.5 indicate that there are indications of clustering in the data, but the strength of the clustering is not very strong.

DOI: https://doi.org/10.17509/seict.v5i2.75942
p-ISSN 2774-1656 | e- ISSN 2774-1699

6. REFERENCES

A. Mustofa, "Pemikiran Harun Yahya Dalam Nilai Nilai Moral Al-Qur'an (Studi Analisis Nilai Nilai Pendidikan Akhlaq)," J. Pendidik. Islam, vol. 4, no. 1, Art. no. 1, 2018

A. Nurrohim and I. N. Sidik, "Ḥikmah Dalam Al-Qur'an: Studi Tematik Terhadap Tafsir Al-Mizān," Profetika J. Studi Islam, vol. 20, no. 2, Art. no. 2, Jan 2020

C.-Y. Chu, K. Park, dan G. E. Kremer, "Applying Text-mining Techniques to Global Supply Chain Region Selection: Considering Regional Differences," Procedia Manuf., vol. 39, hlm. 1691–1698, Jan 2019

D. I. A. Putra and M. Hidayaturrahman, "The roles of technology in al-Quran exegesis in Indonesia," Technol. Soc., vol. 63, hlm. 101418, Nov 2020

D. T. Pham, S. S. Dimov, dan C. D. Nguyen, "Selection of K in K-means clustering," Proc. Inst. Mech. Eng. Part C J. Mech. Eng. Sci., vol. 219, no. 1, hlm. 103–119, Jan 2005

G. Miner, J. E. IV, T. Hill, R. Nisbet, D. Delen, dan A. Fast, Practical Text Mining and Statistical Analysis for Non-structured Text Data Applications, 1st edition. Waltham, MA: Academic Press, 2012.

H. Z. bin H. Thaib, "Tadarus Alquran: Urgensi, Tahapan, dan Penerapannya," Almufida J. Ilmu-Ilmu Keislam., vol. 1, no. 1, Art. no. 1, 2016

H. Hassani, C. Beneki, S. Unger, M. T. Mazinani, dan M. R. Yeganegi, "Text Mining in Big Data Analytics," Big Data Cogn. Comput., vol. 4, no. 1, Art. no. 1, Mar 2020

H. Liang, X. Sun, Y. Sun, dan Y. Gao, "Text feature extraction based on deep learning: a review," EURASIP J. Wirel. Commun. Netw., vol. 2017, no. 1, hlm. 211, Des 2017

H. M. Liu dan J. G. Lu, "Brief Survey of K-Means Clustering Algorithms," Appl. Mech. Mater., vol. 740, hlm. 624–628, 2015

Hendriyana, A. F. Huda, dan Z. A. Baizal, "Feature Extraction Amazon Customer Review to Determine Topic on Smartphone Domain," dalam 2021 13th International Conference on Information & Communication Technology and System (ICTS), 2021, hlm. 342–347

I. Idaman and S. Hidayat, "Al-Qur'an Dan Kecerdasan Spiritual: Upaya Menyingkap Rahasia Allah Dalam Al-Qur'an," Khatulistiwa, vol. 1, no. 1, Mar 2011

I. L. Anggraeni dan Y. A. Rozali, "Quarter Life Crisis Ditinjau Dari Kecerdasan Emosional Pada Dewasa Awal," Psychommunity Semin. Nas. Psikol. Esa Unggul, no. 0, Art. no. 0, 2023, Diakses: 28 Juli 2024.

J. Cao, J. Fang, Z. Meng, dan S. Liang, "Knowledge Graph Embedding: A Survey from the Perspective of Representation Spaces," 15 Oktober 2023

L. Kaufman dan P. Rousseeuw, Finding Groups in Data: An Introduction To Cluster Analysis. 1990.

M. Mardan, Al-Qur'an Sebuah Pengantar. Jakarta: Pustaka Mapan Jakarta, 2010. Diakses: 11 Februari 2024.

M. E. Celebi, H. A. Kingravi, dan P. A. Vela, "A comparative study of efficient initialization methods for the k-means clustering algorithm," Expert Syst. Appl., vol. 40, no. 1, hlm. 200–210, Jan 2013

M. Zahid, "Perbedaan Pendapat Para Ulama Tentang Jumlah Ayat Al-Qur'an Dan Implikasinya Terhadap Penerbitan Mushaf Al-Qur'an Di Indonesia," NUANSA J. Penelit. Ilmu Sos. Dan Keagamaan Islam, vol. 9, no. 1, Art. no. 1, Jan 2012

M. Allahyari dkk., "A Brief Survey of Text Mining: Classification, Clustering and Extraction Techniques," 28 Juli 2017, arXiv: arXiv:1707.02919.

DOI: https://doi.org/10.17509/seict.v5i2.75942
p-ISSN 2774-1656 | e- ISSN 2774-1699

P. Ekman, "An argument for basic emotions," Cogn. Emot., vol. 6, no. 3–4, hlm. 169–200, Mei 1992

R. E. Martin and K. N. Ochsner, "The Neuroscience of Emotion Regulation Development: Implications for Education," Curr. Opin. Behav. Sci., vol. 10, hlm. 142–148, Agu 2016

R. Aritama, "Perkembangan Penerjemahan Al-Quran di Indonesia dari Masa ke Masa," Tafsir Al Quran | Referensi Tafsir di Indonesia. Diakses: 24 Agustus 2024.

S. Suparlan, "Psikologi Dan Kepribadian Perspektif Al-Quran," Humanika Kaji. Ilm. Mata Kuliah Umum, vol. 8, no. 1, Art. no. 1, 2008

S. Sakthi Vel, "Pre-Processing techniques of Text Mining using Computational Linguistics and Python Libraries," dalam 2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS), Mar 2021, hlm. 879–884.

S. Kurniawan, W. Gata, D. A. Puspitawati, I. K. S. Parthama, H. Setiawan, dan S. Hartini, "Text Mining Pre-Processing Using Gata Framework and RapidMiner for Indonesian Sentiment Analysis," IOP Conf. Ser. Mater. Sci. Eng., vol. 835, no. 1, hlm. 012057, Apr 2020

Y. E. Saida, "Pengelompokan terjemahan ayat-ayat Al-Qur'an dalam Bahasa Indonesia dengan Algoritma K-Means Clustering," Sarjana, Universitas Brawijaya, 2007. Diakses: 30 Juli 2024.

DOI: https://doi.org/10.17509/seict.v5i2.75942
p-ISSN 2774-1656 | e- ISSN 2774-1699