Accredited Ranking SINTA 2
Decree of the Director General of Higher Education, Research, and Technology, No. 158/E/KPT/2021
Validity period from Volume 5 Number 2 of 2021 to Volume 10 Number 1 of 2026

Published online on: **http://jurnal.iaii.or.id**

JURNAL RESTI
(Rekayasa Sistem dan Teknologi Informasi)
Vol. 6 No. 5 (2022) 803 - 809 ISSN Media Electronic: 2580-0760

Topic Classification of Quranic Verses in English Translation Using Word Centrality Measurement

Achmad Salim Aiman[1], Kemas Muslim Lhaksmana[2], Jondri[3]
[1,2,3]School of Computing, Telkom University
[1]salimaiman@student.telkomuniversity.ac.id, [2]kemasmuslim@telkomuniversity.ac.id, [3]jondri@telkomuniversity.ac.id

Abstract

Every Muslim in the world believes that the Quran is a miracle and the words of God (Kalamullah) revealed to the Prophet Muhammad SAW to be conveyed to humans. The Quran is used by humans as a guide in dealing with all problems in every aspect of life. To study the Quran, it is necessary to know what topic is being discussed in every single verse. With the help of technology, the verses of the Quran can be given topics automatically. This task is called multilabel classification where input data can be classified into one or more categories. This research aims to apply the multilabel classification to classify the topics of the Quranic verses in English translation into 10 topics using the Word Centrality measurement as the word weighting value. Then a comparison is made to the 4 classification methods, namely SVM, Naïve Bayes, KNN, and Decision Tree. The result of the centrality measurement shows that the word 'Allah' is the most important or the most central word of the whole document of the Quran with the scenario using stopword removal. Furthermore, the use of word centrality value as term weighting in feature extraction can improve the performance of the classification system.

Keywords: *Quran, Topic classification, Multilabel, Word Centrality, SVM, Naïve Bayes, KNN, Decision Tree*

1. Introduction

The Quran is a book that comes from God and it was revealed to the prophet Muhammad SAW as a guide for human life. Quran is a source of teaching and law that contains 6236 verses which are divided into 114 surahs into 30 chapters. As a *Kalamullah* (words of God), the Quran has a unique sentence structure that does not exist in other scriptures. The Quran also has many special characteristics that have been studied by the Islamic scholars. The studies focused on the specialties of the Quran in several aspects. One of the aspects is the main content as the primary source of the religion of Islam.

To understand the Quran, it is necessary to know the topic discussed in each verse. A single verse can be classified into more than one topic. For example, the last verse of Surah Al-Fatihah can be classified into the category of Pillars of Faith and Forbidden Acts. In machine learning, this task is called Multilabel text classification where the main purpose is to classify a text into one or more categories [1]. In the text classification task, every text is converted into a *Vector Space Model* for feature extraction, so that it can be used by machine learning algorithms for the classification process [2]. Word centrality measurement is one method that has been used as feature extraction in several studies and gives good results [3][4].

There are several studies where word centrality measurements are used for the classification task. The research about graph-based term weighting for text categorization by Fragkiskos [2], discusses how a word can be weighted by calculating its centrality value in a graph. The graph used is a directed graph and the centrality measurement using degree centrality, and closeness centrality. The result obtained is that the classification with degree centrality produces the best accuracy of 96%. Another research proposed by Ishtiaq [3] about graph centrality-based term weighting for SMS Spam detection. The centrality methods used are degree centrality and closeness centrality. And two classification methods are used for the classification process: SVM and Naïve Bayes. The obtained is degree centrality produces best recall of 76% and best precision of 81%.

The study of centrality measurement in the original Quran (Arabic scripture) by Ferdian [4] shows that the word 'الله' (*Allah*) is the word with the highest centrality value based on the centrality measurement using degree

Accepted: 29-07-2022 | Received in revised: 10-09-2022 | Published: 31-10-2022

centrality, closeness centrality, and betweenness centrality in Graph of Word (GoW) with stopword removal. The centrality value is used as Term Weight (TW) combined with IDF (TW-IDF) as feature extraction for the classification system using SVM and Naïve Bayes. The result obtained is the hamming loss score reduced from 0.43 to 0.21 on Naïve Bayes classifier [4].

Furthermore, there are several studies discussed about the topic classification of Quranic verses in English translation. First, the study by Pane [5] about a multilabel classification system using Multinomial Naïve Bayes for classifying Quranic verses into 15 topics. The result from this research is the best hamming loss score the system produced is 0.1247 [5]. In another study by Ulumudin [6], a topic classification system for the Quranic verses in English translation built using KNN to classify 3 topics, and also comparing the use of two feature extraction methods, namely: Weighted TF-IDF and TF-IDF. The result shows that the classification using KNN can produce the best hamming loss score of 0.13537 on k=25 with feature extraction using Weighted TF-IDF [6]. Next, a study by Hidayat [7] compared four ML algorithms, namely: SVM, Naïve Bayes, KNN, and Decision Tree on the topic classification system of Quranic verses in English and Indonesian translation. The result obtained is that the four proposed algorithms can provide the accuracy score above 70%.

Based on the related studies, the use of term-weighting using graph-based centrality measurement for single-label or multilabel classification can improve the performance of the classification systems. Thus, for this research we create the same Graph of Word (GoW) model proposed in [4] that represents the relationship of words in all verses of the Quran in English translation. The GoW is created in two models: GoW with stopword removal and GoW without stopword removal. The graph in the GoW model consists of nodes and edges. Nodes represent words in the verses, while the edges represent the relationship between words, and the edges also represent the relationship between the last word of a verse and the first word of the next verse. We also use three centrality measurement methods namely degree centrality, closeness centrality, and betweenness centrality. The centrality value is used as TW (Term Weight) and combined with IDF for feature extraction.

For the classification stage, we use four classification methods, namely: SVM, Naïve Bayes, KNN, and Decision Tree. The reason we selected these methods is there are several advantages and unique characteristics of the methods. Also, the previous related studies show that these algorithms can give the good results for classification systems, especially in the Quranic verse topic classification.

SVM is a classification method that is suitable for unstructured data and structured data like text and images, and it is successfully applied in many applications [8][9]. Naïve Bayes is an algorithm that implement the probabilistic method based on the Bayes' Theorem that has been successfully applied in many studies [5][10]. KNN is a simple method for classification that works by grouping data points into specific neighbors, this method also gives competitive results in several studies [6][11]. Decision Tree is a fast algorithm for classifying unknown records and gives good results in many studies [12].

There are several goals to be achieved. First, we want to identify the most central word of the Quran in English translation. Second, we want to know the effect of the centrality value as TW (Term Weight) combined with IDF as the feature extraction method for the classification system. Lastly, we want to compare the performance of SVM, Naïve Bayes, KNN, and Decision Tree on the multilabel classification system when the centrality measurement values are applied for the feature extraction.

2. Research Methods

We build the system flow into six stages. The system flow is depicted in Figure 1.

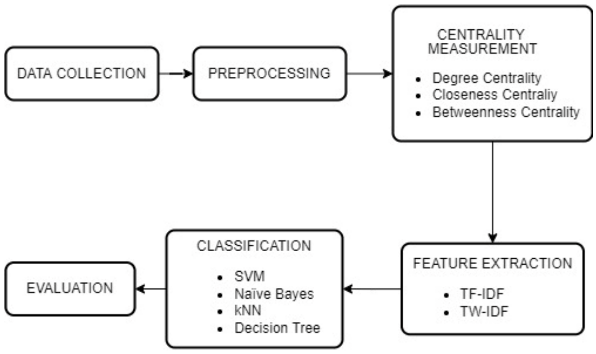

Figure 1. System Flow

2.1 Data Collection

The dataset used for this research is The English Translation of the Quran by Abdullah Yusuf Ali [13] provided and published by a website called Quranyusufali.com. The topic labels are collected using scrapping method on a website called Quranverses.net [14]. There are ten topics provided in [14] that are used for this research, such as Pillars of Faith (1), Mandatory Acts (2), Muslim Ethics (3), Forbidden Acts (4), Foundation of The Laws (5), People of The Book (6), Creation and Creatures (7), Invocation in The Quran (8), Makkah (9), and Science in The Quran (10). The example of the dataset in Surah An-Naas which is the last chapter of the Quran is shown in Table 1.

The total number of data used for building GoW is 6236 verses with 6497 unique words. However, during the

DOI: https://doi.org/10.29207/resti.v6i5.4358

scrapping process, the total data collected are only 4054 labeled verses with 5672 unique words. The remaining 2182 verses are not labeled. This case is happened because there are some verses that are not included in certain labels and some other verses cannot be labeled. For example, the first verse of Surah Al-Baqarah that reads "الم" (*Alif Lam Mim*) cannot be labeled because the verse agreed by the Islamic scholars has no meaning.

Table 1. Surah An-Naas in the dataset

Surah/ Ayah	Text	Topic
114/1	Say: I seek refuge with the Lord and Cherisher of Mankind,	['Pillars of Faith']
114/2	The King (or Ruler) of Mankind,	['Pillars of Faith']
114/3	The god (or judge) of Mankind,	['Pillars of Faith']
114/4	From the mischief of the Whisperer (of Evil), who withdraws (after his whisper),	['Forbidden Acts']
114/5	(The same) who whispers into the hearts of Mankind,	['Forbidden Acts']
114/6	Among Jinns and among men.	['Creation and Creatures', 'Pillars of Faith']

2.2 Data Preprocessing

The data preprocessing stage is caried out to improve the data quality before the data is being processed. Preprocessing consists of three stages, such as text cleaning, tokenization, and stopword removal. Text cleaning is done by transforming all letters into lowercase form, and removing special characters and punctuations. In the tokenization stage, each sentence is separated into words as tokens. After tokens are generated, the next thing we perform is stopword removal. Stopword removal is a stage where words that often appear and do not represent the meaning of a sentence are removed. Stopword removal on Quranic verses is not intended to make any change on the content of the Quran which is prohibited in Islam[4]. The example of input and output in each preprocessing stage can be seen in Table 2.

2.3 Centrality Measurement

For the centrality measurement, first we have to build a Graph of Word (GoW). A GoW is a graph model that represents the connection between words in the Quranic verse. In GoW, nodes represent every single word from the whole document (Quran). Nodes are interconnected by edges. The edges represent the co-occurrence relationship between words in a verse, also, edges represent the relation between the last word of a verse and the first word of the next verse. The weights on the edge represent the number of co-occurrence relations between words. The weight of an edge will increase if there are multiple relations occurring between a pair of nodes.

Table 2. Example of input and output of Data Preprocessing

Stages	Input	Output
Text Cleaning	Among Jinns and among men.	among jinns and among men
Tokenization	among jinns and among men	['among', 'jinns', 'and', 'among', 'men']
Stopword Removal	['among', 'jinns', 'and', 'among', 'men']	['among', 'jinns', 'among', 'men']

In this research, we built a GoW in an Undirected graph model. In Figure 2 we can see an example of GoW built from Surah Al-Ikhlas with four verses. We can see that the weight of the edge that connects the node 'He' and node 'is' is 2. This is because the co-occurrence relation of the word 'He' and 'is' appears two times, in the first and the third verse of the Surah. The first verse reads "*Say **He is** Allah the one and only.*", and the third verse reads "*He begetteth not, nor **is He** begotten.*".

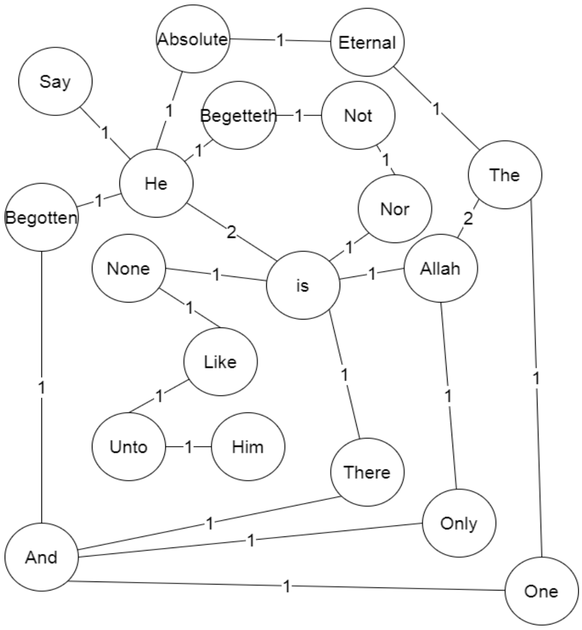

Figure 2. Example of GoW in Surah Al-Ikhlas

We built two models of the graph, first GoW with stopword removal and second GoW without stopword removal. Before building the GoW model we have to preprocess the text data by performing text cleaning and tokenization. Then for the first GoW model, we perform stopword removal to the preprocessed text data. The process of building GoW can be seen in Figure 3.

After building the GoW models, then we measure the word centrality using three centrality methods such as degree centrality, closeness centrality, and betweenness centrality. Based on the graph theory, centrality measurement is used to determine the most important (central) nodes in a particular graph. In our case, we use

DOI: https://doi.org/10.29207/resti.v6i5.4358

the centrality measurement to find the most important word in the Quran.

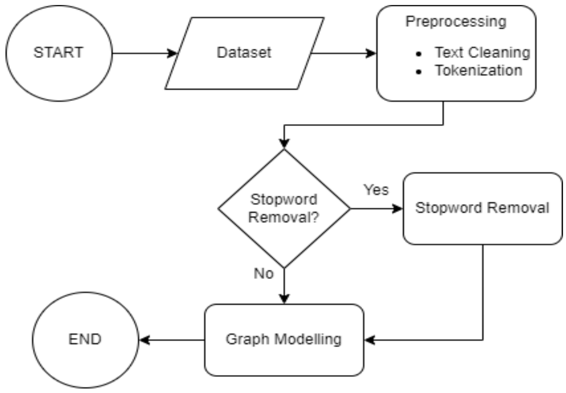

Figure 3. Graph Modeling flow

2.3.1 Degree Centrality

Degree centrality is measured by calculating the number of edges that are connected to a certain node. In other words, degree centrality is defined as the number of connections owned by a node. In GoW model, degree centrality value is equal to the number of words (nodes) that has a connection with node *i* [15]. Every node *j* will have a connection with node *i* if the word of node *j* appears before or after the word of node *i*. Degree centrality can be calculated using the formula in equation (1).

$$C_{Deg}(i) = \frac{|N(i)|}{|V|-1} \quad (1)$$

Where $C_{Deg}(i)$ is the value of degree centrality of node *i*, *N(i)* represents the number of nodes connected with node *i*, and *V* represents the total amount of nodes in the GoW [4].

2.3.2 Closeness Centrality

Closeness centrality is defined as how close is a node to another node[16]. This method aims to calculate the distance between nodes. The closeness centrality can be calculated using the formula in equation (2).

$$C_{Close}(i) = \frac{N-1}{\sum_{u \in j} d(i,j)} \quad (2)$$

Where $C_{Close}(i)$ is the closeness centrality value of node *i*, *d(i,j)* is the shortest path of node *i* to *j*, and *N* is the total number of nodes in the GoW [4].

2.3.3 Betweenness Centrality

Betweenness centrality is calculated by the total amount of times a node is being a connecting line between a set of nodes to another set of nodes [16]. Betweenness centrality can be calculated using the formula in equation (3).

$$C_{Bet}(i) = \sum_{j<k} {g_{jk}(i)}/{g_{jk}} \quad (3)$$

Where $C_{Bet}(i)$ is the betweenness centrality value of node *i*, g_{jk} is the number of shortest paths connecting node *j* and *k*, while $g_{jk}(i)$ is the number of shortest paths connecting node *j* and *k* through node *i*.

2.4 Feature Extraction

Feature extraction is used to give some weights to a word. The text data is converted into a numerical form, so the data can be processed by the machine learning algorithms [17]. The text data is represented in a *vector space model* and weighted using several methods such as Bag-of-Words (BoW) and TF-IDF.

TF-IDF is one example of a feature extraction method that is used widely. TF-IDF is calculated by combining the TF (Term Frequency) with IDF (Inverse Document Frequency) [17]. TF is the frequency of occurrence of a word, while IDF is calculated by dividing the total number of documents by the number of documents that contain a particular word and calculating the logarithm. [17]. The calculation of TF-IDF can be done using the formula in equation (4).

$$w_{t,d} = tf_{t,d} \times log\left(\frac{D}{df_t}\right) \quad (4)$$

Where $w_{t,c}$ is the TF-IDF weight of term *t* in document *d*, $tf_{t,d}$ is the total occurrence of term *t* in document *d*, *D* is the total number of documents, and df_t is the number of document that contains term *t* [4].

The word centrality values that have been calculated in the centrality measurement stage using formulas in equation (1), (2), and (3), can also be applied for feature extraction. The centrality values can be implemented as TW (Term Weight) and combined with IDF as TW-IDF. We can calculate TW-IDF using the formula in equation (5).

$$w_{t,d} = tw_t \times log\left(\frac{D}{df_t}\right) \quad (5)$$

Where $w_{t,d}$ is the TW-IDF weight of term *t* in document *d*, and tw_t is the centrality value of term *t*.

2.5 Classification

The system model is built using four classification methods to see the effect of using word centrality. We used four classification methods, namely: SVM, Naïve Bayes, KNN, and Decision Tree for classifying 10 topics defined in Data Collection stage.

2.5.1 Support Vector Machine (SVM)

SVM is a powerful method for the single-label, multilabel, and multiclass classification tasks. SVM is a supervised learning method that works by dividing data into several classes by determining the maximum margin of the hyperplane [18]. This method works in several phases. First, SVM will search for the support vector in every single class.

Table 3. Top 10 Word with Highest Centrality Value based on Degree Centrality with Stopword Removal

Word	Centrality
Allah	0.1843
Ye	0.1542
Lord	0.0818
Say	0.0767
Said	0.0756
One	0.0719
Thee	0.0689
Thou	0.0656
Shall	0.0644
Us	0.0610

Table 4. Top 10 Word with Highest Centrality Value based on Closeness Centrality with Stopword Removal

Word	Centrality
Allah	0.5224
Ye	0.5104
Lord	0.4753
Say	0.4741
Said	0.4719
One	0.4663
Shall	0.4644
Thou	0.4632
Us	0.4617
Thee	0.4607

A support vector is a sample data point from each class that has the closest distance to the other classes. After the support vectors are obtained, then the algorithm calculates the margin to determine highest margin [18].

2.5.2 Naïve Bayes

Naïve Bayes is a machine learning model that implements probabilistic approach that is used for classification task [10]. Naïve Bayes is known as a probabilistic algorithm that implements the Bayes' Theorem where the algorithm makes a predictive analysis based on the probability of a data. Naïve Bayes is a simple algorithm and easy-to-apply method and it is used for classification tasks widely [19].

2.5.3 K-Nearest Neighbors

KNN is a simple classification algorithm that works by calculating the minimum distance as a parameter to measure the similarity of data for the classification process. KNN is a supervised learning method. KNN is also called a lazy learning method because in the classification process this method uses another data object to classify the new data object.

This method works by determining the *k* number of data points that have the minimum distance to the new data point [6].

Table 5. Top 10 Word with Highest Centrality Value based on Betweenness Centrality with Stopword Removal

Word	Centrality
Allah	0.1563
Ye	0.1213
Said	0.0436
Lord	0.0409
One	0.0397
Say	0.0374
Thee	0.0345
Thou	0.0327
Shall	0.0310
Day	0.0258

2.5.4 Decision Tree

DT works by building the classification trees to make a decision for data classification. This method forms the branches which are decision-making steps that can lead to a clear classification result [12].

2.6 Evaluation

In this research we use hamming loss method for evaluating the performance of the classification method. The hamming loss method is suitable for multilabel classification tasks because the hamming loss works by calculating prediction error and missing error [6]. The calculation of Hamming loss method can be done using the formula in equation (6).

$$HL = \frac{1}{NL}\sum_{i=1}^{N}\sum_{j=1}^{L}\left[\hat{y}_j^{(i)} \neq y_j^{(i)}\right] \quad (6)$$

Where N is the total documents, L is the total classes, $\hat{y}_j^{(i)}$ is the number of predicted data, $y_j^{(i)}$ is the number of actual data.

3. Results and Discussions

In the experiment process, we determined the most important word of the Quran by performing the word centrality measurement. Then we study the effect on the use of stopword removal on feature extraction applied to the classification system. We also study the comparison of the performance of the classifiers.

3.1 Quranic Word Centrality

The centrality measurement of the Quranic word has been done in two GoW models: GoW with stopword removal and GoW without stopword removal. The result of the centrality measurement can be seen in Table 3, 4, 5, 6, 7, and 8.

Table 6. Top 10 Word with Highest Centrality Value based on Degree Centrality without Stopword Removal

Word	Centrality
And	0.3075
The	0.2597
Of	0.2168
To	0.1955
In	0.1644
For	0.1263
They	0.1234
But	0.1116
That	0.1047
Their	0.1046

The tables show the ten words with the highest centrality values based on every centrality measurement method with and without using stopword removal.

DOI: https://doi.org/10.29207/resti.v6i5.4358

Table 7. Top 10 Word with Highest Centrality Value based on Closeness Centrality without Stopword Removal

Word	Centrality
And	0.5866
The	0.5681
Of	0.5415
To	0.5380
In	0.5341
For	0.5215
But	0.5208
They	0.5188
That	0.5173
It	0.5155

In the GoW using stopword removal showed in Table 3, 4, and 5, we can see that the word '*Allah*' (God) is the most important word with the centrality values of 0.1843, 0.5224, and 0.1563 on degree centrality, closeness centrality, and betweenness centrality.

Table 6, 7, and 8 show the GoW model without stopword removal, every word in the tables are stopwords. It means that all top-10 words are stopword. The word 'And' has the highest degree centrality, closeness centrality, and betweenness centrality value at 0.3075, 0.5866, and 0.1711.

In the previous research about centrality measurement in the original Quran (Arabic) [4], the word "الله" (Allah) is the second most central or important word in GoW without stopword removal, and the most central or important word in GoW with stopword removal [4]. Thus, the word "الله" (Allah) is always in the top-two of the most central words [4].

Table 8. Top 10 Word with Highest Centrality Value based on Betweenness Centrality without Stopword Removal

Word	Centrality
And	0.1711
The	0.1321
Of	0.1070
To	0.0797
In	0.0619
For	0.0397
They	0.0358
That	0.0320
Them	0.0310
He	0.0298

If we compare that result to our result, it must be admitted that the Arabic language of the Quran is more meaningful than the English translation. A word in Arabic can have more than one meaning, and also, a word can be interpreted with more than one translation word. This can lead to more and more stopwords, so that it can affect the centrality measurement process.

3.2 Topic Classification of Quranic Verses

The classification process is done by applying four machine learning methods, namely: SVM, Naïve Bayes, KNN, and Decision Tree. Comparisons were also made on four feature extraction methods applied to the classification system, such as TF-IDF, TW-IDF Deg (degree centrality), TW-IDF Close (closeness centrality), and TW-IDF Bet (betweenness centrality). Also, a comparison of the GoW with stopword removal and without stopword removal applied to the classification system. Then, the system performance is evaluated using the hamming loss. The result of the classification processes can be seen in Table 9 and Table 10.

The effect of using stopword removal can be concluded by comparing both tables. It is clear that the four classification methods can be slightly improved with the usage of stopword removal. It can be seen that the difference between each hamming loss score in both tables is not much. Therefore, we can conclude that the use of stopword removal has no significant effect for the system performance. However, overall, the use of stopword removal gives better results for the system.

The word centrality values that have been obtained that described in the previous subsection are applied as TW (Term Weight) value which are then used in the feature extraction stage. In Table 9, it can be seen that feature extraction using TW-IDF Close (Closeness Centrality) can improve the performance of three classification systems, namely Naïve Bayes, KNN, and Decision Tree.

Table 9. Result of the scenario With Stopword Removal

Machine Learning	Feature Extraction			
	TD-IDF	TW-IDF *Deg*	TW-IDF *Close*	TW-IDF *Bet*
SVM	**0.1110**	0.1292	0.1199	0.1413
NB	0.1252	0.1467	**0.1224**	0.1498
KNN	0.1287	0.1459	**0.1224**	0.1464
DT	0.1325	0.1249	**0.1247**	0.1264

In Table 10, the use of the word centrality value can also slightly improve the performance of SVM on TW-IDF Close and TW-IDF Bet, and the performance of KNN on TW-IDF Deg and TW-IDF Bet. For the SVM model in Table 9, as well as the SVM, Naïve Bayes, and KNN models in Table 10, feature extraction using TF-IDF gives higher effect on improving the performance of the classification systems. However, the feature extraction using TW-IDF can still improve the performance of the classification system.

Lastly, to compare the four classification methods, it is done by comparing the hamming loss scores of each classifier. In both tables, SVM produces the lowest (best) hamming loss score of 0.1110 and the highest (worst) hamming loss score of 0.1413. Naïve Bayes results the hamming loss score between 0.1221 and 0.1498. Also, the hamming loss score of KNN is between 0.1224 and 0.1464. Finally, the hamming loss score of Decision Tree is between 0.1247 and 0.1412. Thus, we can conclude that SVM outperforms Naïve Bayes, KNN, and Decision Tree in the topic classification of the Quranic verses in English translation.

DOI: https://doi.org/10.29207/resti.v6i5.4358

4. Conclusion

The goal of this study is to determine the effect of the use of centrality measurement of Quranic words for the classification of Quranic topics in English translation, where the centrality value is used as term-weighting for feature extraction. Then, four machine learning methods (SVM, Naïve Bayes, KNN, and Decision Tree) are used to evaluate the effect of word centrality measurement on the performance of the classification system. After getting the results of this research, we can conclude that the word centrality measurement can improve the performance of the classification system, especially when using the closeness centrality method and by applying stopword removal, which can produce the best hamming loss scores on Naïve Bayes, KNN, and Decision Tree.

For future works, we consider using more datasets and use more topics. Also, we can use the different datasets such as The Quran in other languages, and consider to apply other centrality methods such as Eccentricity Centrality and Eigenvector Centrality.

Table 10. Result of the scenario Without Stopword Removal

Machine Learning	Feature Extraction			
	TF-IDF	TW-IDF *Deg*	TW-IDF *Close*	TW-IDF *Bet*
SVM	**0.1133**	0.1334	0.1165	0.1401
NB	**0.1221**	0.1474	0.1267	0.1498
KNN	**0.1252**	0.1443	0.1447	0.1441
DT	0.1412	**0.1299**	0.1324	0.1318

Reference

[1] A. Pal, M. Selvakumar, and M. Sankarasubbu, "Multi-Label Text Classification using Attention-based Graph Neural Network," Mar. 2020, doi: 10.5220/0008940304940505.

[2] F. D. Malliaros and K. Skianis, "Graph-Based Term Weighting for Text Categorization," in *Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2015*, Aug. 2015, pp. 1473–1479. doi: 10.1145/2808797.2808872.

[3] A. Ishtiaq, M. A. Islam, M. Azhar Iqbal, M. Aleem, and U. Ahmed, "Graph Centrality Based Spam SMS Detection," in *2019 16th International Bhurban Conference on Applied Sciences and Technology (IBCAST)*, Jan. 2019, pp. 629–633. doi: 10.1109/IBCAST.2019.8667174.

[4] F. Yulianto, K. M. Lhaksmana, and D. T. Murdiansyah, "Classifying Quranic Verse Topics using Word Centrality Measure," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 5, no. 3, pp. 594–601, Jun. 2021, doi: 10.29207/resti.v5i3.3171.

[5] R. A. Pane, M. S. Mubarok, N. S. Huda, and Adiwijaya, "A Multi-Lable Classification on Topics of Quranic Verses in English Translation Using Multinomial Naive Bayes," in *2018 6th International Conference on Information and Communication Technology (ICoICT)*, May 2018, pp. 481–484. doi: 10.1109/ICoICT.2018.8528777.

[6] G. I. Ulumudin, A. Adiwijaya, and M. S. Mubarok, "A multilabel classification on topics of qur'anic verses in English translation using K-Nearest Neighbor method with Weighted TF-IDF," *J Phys Conf Ser*, vol. 1192, p. 012026, Mar. 2019, doi: 10.1088/1742-6596/1192/1/012026.

[7] R. Hidayat and S. Minati, "Comparative Analysis of Text Mining Classification Algorithms for English and Indonesian Qur'an Translation," *IJID (International Journal on Informatics for Development)*, vol. 8, no. 1, p. 47, Jun. 2019, doi: 10.14421/ijid.2019.08108.

[8] D. A. Otchere, T. O. Arbi Ganat, R. Gholami, and S. Ridha, "Application of supervised machine learning paradigms in the prediction of petroleum reservoir properties: Comparative analysis of ANN and SVM models," *J Pet Sci Eng*, vol. 200, p. 108182, May 2021, doi: 10.1016/j.petrol.2020.108182.

[9] K. S. Sahoo *et al.*, "An Evolutionary SVM Model for DDOS Attack Detection in Software Defined Networks," *IEEE Access*, vol. 8, pp. 132502–132513, 2020, doi: 10.1109/ACCESS.2020.3009733.

[10] M. R. Choirulfikri, K. M. Lhaksamana, and S. al Faraby, "A Multi-Label Classification of Al-Quran Verses Using Ensemble Method and Naïve Bayes," *Building of Informatics, Technology and Science (BITS)*, vol. 3, no. 4, pp. 473–479, Mar. 2022, doi: 10.47065/bits.v3i4.1287.

[11] H. Pratama, "Machine Learning: Using Optimized KNN (K-Nearest Neighbors) to Predict the Facies Classifications," in *Proceedings of the 13th SEGJ International Symposium, Tokyo, Japan, 12–14 November 2018*, Apr. 2019, pp. 538–541. doi: 10.1190/SEGJ2018-139.1.

[12] S. N. Safitri, Haryono Setiadi, and E. Suryani, "Educational Data Mining Using Cluster Analysis Methods and Decision Trees based on Log Mining," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 6, no. 3, pp. 448–456, Jul. 2022, doi: 10.29207/resti.v6i3.3935.

[13] "The Holy Quran Translation by Abdullah Yusuf Ali, 2022." https://quranyusufali.com/ (accessed Jul. 29, 2022).

[14] "Quran Verses – Verses from the Noble Qur'an categorized according to subject and accompanied by a beautiful recitation." https://quranverses.net/ (accessed Jul. 29, 2022).

[15] S. Kr. Biswas, M. Bordoloi, and J. Shreya, "A graph based keyword extraction model using collective node weight," *Expert Syst Appl*, vol. 97, pp. 51–59, May 2018, doi: 10.1016/j.eswa.2017.12.025.

[16] J. Zhang and Y. Luo, "Degree Centrality, Betweenness Centrality, and Closeness Centrality in Social Network," 2017. doi: 10.2991/msam-17.2017.68.

[17] D. Kim, D. Seo, S. Cho, and P. Kang, "Multi-co-training for document classification using various document representations: TF–IDF, LDA, and Doc2Vec," *Inf Sci (N Y)*, vol. 477, pp. 15–29, Mar. 2019, doi: 10.1016/j.ins.2018.10.006.

[18] M. Goudjil, M. Koudil, M. Bedda, and N. Ghoggali, "A Novel Active Learning Method Using SVM for Text Classification," *International Journal of Automation and Computing*, vol. 15, no. 3, pp. 290–298, Jun. 2018, doi: 10.1007/s11633-015-0912-z.

[19] V. K. Chauhan, K. Dahiya, and A. Sharma, "Problem formulations and solvers in linear SVM: a review," *Artif Intell Rev*, vol. 52, no. 2, pp. 803–855, Aug. 2019, doi: 10.1007/s10462-018-9614-6.

DOI: https://doi.org/10.29207/resti.v6i5.4358