20220611

by Syarif Syarif

Submission date: 11-Jun-2022 01:15PM (UTC+0800)
Submission ID: 1854696079
File name: 2356-Article_Text-7267-1-10-20220604.pdf (1.14M)
Word count: 4294
Character count: 21724

Jurnal Mantik, 6 (1) (2022) 639-648

Published by:Institute of Computer Science (IOCS)

Jurnal Mantik

Journal homepage: www.iocscience.org/ejournal/index.php/mantik/index

Classification of Al-Qur'an Arabic Verses Used Naive Bayes

Syarif Hidayatullah

Informatics Tekhnical
University of Muhammadiyah Jember, Karimata Street, No. 49, Jember, 6812, Indonesia

E-mail: syarifhidayatullah@unmuhjember.ac.id

ARTICLEINFO

Article history:

Received: Apr 13, 2022
Revised: May 10, 2022
Accepted: May 29, 2022

Keywords:

Al-Quran,
Classification,
Naïve Bayes,
Text Mining.

ABSTRACT

The development of information globally is very fast, it demands the provision of media information can be enjoyed and felt quickly and precisely. Al-Qur'an as a medium of information in the classical form and contains of science, life, etc. that must be distributed to all human beings orally, in writing, and daily behavior, especially in the field of Islamic laws and aspects of social procedures in Islam. This study applies a classification technique using the method Naïve Bayes which there are three classes or categories namely Sholat (Prayer), Hajj and Wedding. The grouping of these verses takes data from the book of LubaabutTafsir Min Ibn Kathir which later obtained accuracy results 75.52%.

1. Introduction

The development of information globally is very rapid, demands the provision of such information media can be enjoyed and felt quickly and precisely [1]. So it is necessary in the aspect of information media must comply with existing norms and procedures. Al-Qur'an as a medium of information in classical form but contains knowledge, life, etc.must be distributed to all human beings orally, in writing, and in daily behavior, especially in the field of Islamic laws and aspects of social procedures in Islam.

Al-Qur'an Al-Karim introduces himself with various features and characteristics. One of them is that Al-Qur'an is a book whose authenticity is guaranteed by Allah, and it is a book that's always maintained [2]. Inna nahnunazzalna al-dhikrawainnalahulahafizhun (Indeed, We sent down the Qur'an and We are the Preservers of Him) (Surah Hijr: 9) [3]. Thus God in guarantee of the authenticity of the Qur'an, the guarantee given is on the basis of His omnipotence and majesty, and thanks to the efforts of His creatures, especially the beloved of Allah, which consists of friends, tabi'in, tabi'ittabi'in, and their successors, namely Ulama [2]. One of the virtues of reading the Qur'an is that it will be elevated by Allah SWT, asthe Prophet Muhammad SAW has been said, namely: Umar bin Khattabra said that the Messenger of Allah had said, "Verily Allah will raise people with this book (the Qur'an), with it Allah will also bear the other people" (HR Muslim) [4]. That Al-Qur'an which in it there are verses of the word of Allah SWT is a source of reference for Muslims both in the field of worship, social, political, mu'amalah (buying and selling transactions) etc..

In Al-Qur'an there are 30 juzes and 114 surahs and 6666 verses, and in the verses there are several sources of reference for Muslims both in the field of monotheism (divinity), wisdom of example, science, Islamic law etc.. Of the many problems of Muslims which refers to the verses of Al-Qur'an, this research focuses only on taking the basics of Islam contained in the verses of the Qur'an. In which there are several verses about marriage, sholat, mu'amalah, hajj, fasting and many other verses contained in Al-Qur'an.

In this study, we will discuss the classification of Arabic texts, especially in the verses of Al-Qur'an, for the text classification methods that have been used in research, such as N-gram [5], k nearest neighbor (k-NN) [6], Decision Tree [7], Naïve Bayes (NB) [8], Support Vector Machine (SVM) [9] Maximum Entropy (ME) [10], these are several methods that have been proposed in text classification.

Below it is a summary of several related studies on the classification of Arabic texts namely Naïve Bayes By El-Kourdi [11] for automatic classification of Arabic text documents. In this study, it produced an

Accredited "Rank 4"(Sinta 4), DIKTI, No. 36/E/KPT/2019, December 13th 2019.

average accuracy of 68.78% and the best accuracy yields 92.8%. In this case El-Kourdi used a number of 1500 text documents which have 5 categories. Every 300 text documents has one category. all the words in the document are converted to the text root of each document. The word size of the sum result is 2,000 conditions/roots. cross-validation was used for evaluation.

Maximum Entropy (ME) is used by El-Halees [10] for Arabic text classification and Sawaf et. Al [12] (2001) to classify and cluster the news articles where the best accuracy reported by El-Halees was 80.4% and 62.7% obtained by El Sawaf.

Association Rules are used by El-Halees [13], and Al-Zoghby [14] to classify the document of Arab. The accuracy reported by Al-Halees shows 74.41%. Al-Zoghbyused the CHARM algorithm and demonstrated its superiority over soft-matching was larger O exact matching. Al-Zoghby used a corpus consisting of 5524 records. Each record is a snippet of the email that has the subject of "Nuclear". The vocabulary size after removing stopwords and punctuation marks is 103,253 words. The average document text size is 18 words. Some words from the document text that have been converted to the root form.

Harrag et. al. [7] improved the classification of Arabic text with the selection of Hybrid features Approach. Harrag uses the Algorima Decesion tree and the classification reports accuracy from 93% for the scientific corpus and 91% for the literary corpus. Harrag collects 2 corpora: the the first is from the scientific insklopedia "Do You Know" (هل تعلم) (it contains 373 documents which has 1 to 8 categories (innovation, geography, sports, famous people, religion, history, Human bodies, and cosmology), from each category have 35 documents. second corpus derived from the hadith insklopedia (موسوعة الحديث الشريف) (from 9 books . التسعة الكتبini contains 435 documents It has 14 categories.

KNN has been implemented by Al-Shalabi et. Al. [6] in arabic texts, they use tf-idf as a scheme for weighting and gaining 95% accuracy. They also apply stemming and feature selection. The authors report in their paper a free issue public shortage for the availability of arabic corpus. They collected an Arabic corpus from several newspapers (Al-Jazeera, An-Nahar, Al-Hayat, Al-Ahram, and Ad-Dostor) and from the website Agriculture Organization. The corpus consists of 621 documents that have 1 to 6 categories. 9 (politics 111, economics 179, exercise 96, health and medicine 114, health and cancer 27, agriculture 100). They have implemented corpus with the removal of stopwords and light stemming.

Leila Kheisat [15] used N-gram frequency statistics in classifying Arabic texts. The writer showed the high dimensional text data with mapping document text in setting real numbers which represents the tri-gram frequency profile. The N-gram method is independent language and works well with the noisy-text case. Tri-grams for the word (المسافر) adalah (.افر, ساف ، مسا ، لمس ، الم). Kheirsat classifies the textdocumentswith a test to calculate Manhattan / Dice similarity distance measures for all training documents and sets the class of the training document to the smallest / distance calculated largest to test the text documents. Kheirat reports that Dice outperforms Manhattan in terms of measure distance. Although Manhattan has provided good classification measures for English text document, it does not seem suitable for Arabic text document. Kheirat collects His corpus of Jordanian newspapers (Al-Arab, Al-Ghad, AlRa'I, Ad-Dostor). It has 1 to 4 corpus categories (sports, economy, weather, and technology). It applied stopwords removal and use 40% for training and 60% for testing.

Faozi H. and El-Qawasmah[16] have used Neural Network (NN) for classification Arabic text. The results of their experiments, showed that in the use of NN with Singular Value Decomposition (SVD) as a feature technique of choice provides better results i.e. (88.3%) compared to basic NN (without SVD) (85.7%) they also experienced high scalability with dimension text datasets using NN. Harrag collects his corpus from Enklopedia hadith from 9 books. It contains 435 documents that have 14 category. He applied light stemming and stopwords removal on his corpus. tf-idf for used as a weighting scheme.

Naïve Bayes has a high level of accuracy to classify Arabic text [17]. Some of the advantages of Naïve Bayes method are the simplicity in computation, it also has several shortcomings in assumptions that are very difficult to fulfill, namely the independence of word features [18]. The Naïve Bayes method in classifying Arabic texts only focuses on general language, specifically this research will be applied in classifying the Arabic text of the verses of Al-Qur'an in categorizing several verses about Islam taken from several existing verses, for the verses related to this research there are only three categories of verses, namely about Sholat, Hajj and Marriage where some of the verses in the categorization are taken from the Lubaabut Tafsir Min Ibn Kathir.

In this study, Naïve Bayes will be applied to classify the Arabic text of the verses of Al-Qur'an to classify several verses contained in the Qur'an with accurate results.

Accredited "Rank 4"(Sinta 4), DIKTI, No. 36/E/KPT/2019, December 13th 2019.

2. Method

2.1 Types of research

The type of research carried out in this research is the experimental method, namely to determine the effect of the Naïve Bayes method on the classification of the Arabic text of the verses of Al-Qur'an in text categorization with the results of the Recall, Precision, and Accuracy values which are implemented in testing text data which is then followed by the implementation on rapidminer and see the level of Recall, Precision, and Accuracy values which are better.

2.2 Method of collecting data

Based on the source, the data of this study were obtained from secondary data which obtained from existing sources, for example the data obtained from sources of the Book, Hadith, etc. For this research, the data obtained is the data of the verses of Al-Qur'an, which is taken from the website *www.textminingthequran.com* in text format that has been sorted according to the number of verses in Al-Qur'an so that it is easy to import it into the database for continuation research, and the categorization of the verse is taken from the classic book, namely Lubaabut Tafsir Min Ibn Kathir as for the categories that have been selected, namely:

Table 1
Categories of Selected Qur'anic Verses

Category
Sholat Verse
Hajj Verse
Marriage verse

As for the categorization, the author quoted from several sources of the yellow book, that is the book of fiqh in which there are already special chapters that discuss this category.

2.3 Proposed Settlement Method

In classifying the Arabic text of the verses of Al-Qur'an, it is necessary to the proposed method then it is used to classify the verses of Al-Qur'an. The method used in this research is the Naïve Bayes Classification (NBC). In the formation of the classification of the Arabic text, it is intended to determine the content of the verses of the Qur'an and to know the accuracy of the verses according to the categories above.

2.4 Experiment and Test Method

Before determining thedata test, that is in the form of verses of Al-Qur'an that have been preprocessed. The form of the process can be described as shown below.

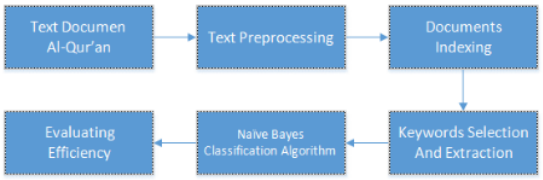

Fig 1: Method Testing Process

Some of these processes can be concluded, namely;

a. Document text is Al-Qur'an that will be classified.
b. Preprocessing text where the basic words are taken by using the stemming process algorithm after preprocessing text that is stage
c. Document Indexing where in one document there are a number of different categories of text after that
d. Keyword selection and extraction in this process will select several keywords for sorting out the verses of Al-Qur'an.
e. Categorization Algorithm or classification of verse groups according to the category of verses and final
f. Evaluation of the text accuracy in determining the Islamic laws that have been determined

Accredited "Rank 4" (Sinta 4), DIKTI, No. 36/E/KPT/2019, December 13th 2019.

2.5 Evaluation and Validation Results

To see the success of an algorithm in classifying Arabic text, the desired verses of Al-Qur'an are to look at the results of testing and training to see the good accuracy level.

3. Results and Discussion

3.1 Data Preparation

In this study, the data used were sourced from Al-Quran which was obtained from the website *www.qurandatabase.org* in the form of xlsx or excel format and also Al-Quran without standard (harkat) sourced from *www.textminingthealquran.com*. which in this research will discuss the classification of the verses of Al-Qur'an which consists of 114 surahs and 6666 verses in the whole. In this study, what will be discussed is the issue of Sholat, Hajj and Marriage. The process is by sorting out the verses of Al-Qur'anand from these verses, several stages will be carried out, namely prepossessing the text of the Qur'anic document, followed by classification process and continued with the evaluation of the series of research processes. Below this is a document of Al-Qur'an before the prepossess of the text. The tables below are part of the texts that will be discussed to carry out the classification process. As for the stages the evaluation will use the overall data.

Table 2
Al-Qur'an Document

Document	Verse	Class
D1	يٰٓاَيُّهَا النَّاسُ اتَّقُوْا رَبَّكُمُ الَّذِيْ خَلَقَكُمْ مِّنْ نَّفْسٍ وَّاحِدَةٍ وَّخَلَقَ مِنْهَا زَوْجَهَا وَبَثَّ مِنْهُمَا رِجَالًا كَثِيْرًا وَّنِسَاۤءً ۚ وَاتَّقُوا اللّٰهَ الَّذِيْ تَسَاۤءَلُوْنَ بِهٖ وَالْاَرْحَامَ ۗ اِنَّ اللّٰهَ كَانَ عَلَيْكُمْ رَقِيْبًا	Merriage
D2	وَاَقِيْمُوا الصَّلٰوةَ وَاٰتُوا الزَّكٰوةَ وَارْكَعُوْا مَعَ الرّٰكِعِيْنَ	Sholat
D3	وَاَقِيْمُوا الصَّلٰوةَ وَاٰتُوا الزَّكٰوةَ وَارْكَعُوْا مَعَ الرّٰكِعِيْنَ	Sholat
D4	يٰٓاَيُّهَا الَّذِيْنَ اٰمَنُوْا لَا تُحِلُّوْا شَعَاۤىِٕرَ اللّٰهِ وَلَا الشَّهْرَ الْحَرَامَ وَلَا الْهَدْيَ وَلَا الْقَلَاۤىِٕدَ وَلَآ اٰۤمِّيْنَ الْبَيْتَ الْحَرَامَ يَبْتَغُوْنَ فَضْلًا مِّنْ رَّبِّهِمْ وَرِضْوَانًا ۗ وَاِذَا حَلَلْتُمْ فَاصْطَادُوْا ۗ وَلَا يَجْرِمَنَّكُمْ شَنَاٰنُ قَوْمٍ اَنْ صَدُّوْكُمْ عَنِ الْمَسْجِدِ الْحَرَامِ اَنْ تَعْتَدُوْا ۘ وَتَعَاوَنُوْا عَلَى الْبِرِّ وَالتَّقْوٰى ۖ وَلَا تَعَاوَنُوْا عَلَى الْاِثْمِ وَالْعُدْوَانِ ۚ وَاتَّقُوا اللّٰهَ ۗ اِنَّ اللّٰهَ شَدِيْدُ الْعِقَابِ	Hajj
D5	وَمِنْ اٰيٰتِهٖٓ اَنْ خَلَقَ لَكُمْ مِّنْ اَنْفُسِكُمْ اَزْوَاجًا لِّتَسْكُنُوْٓا اِلَيْهَا وَجَعَلَ بَيْنَكُمْ مَّوَدَّةً وَّرَحْمَةً ۗ اِنَّ فِيْ ذٰلِكَ لَاٰيٰتٍ لِّقَوْمٍ يَّتَفَكَّرُوْنَ	Merriage
D6	وَاَنْكِحُوا الْاَيَامٰى مِنْكُمْ وَالصّٰلِحِيْنَ مِنْ عِبَادِكُمْ وَاِمَاۤىِٕكُمْ ۗ اِنْ يَّكُوْنُوْا فُقَرَاۤءَ يُغْنِهِمُ اللّٰهُ مِنْ فَضْلِهٖ ۗ وَاللّٰهُ وَاسِعٌ عَلِيْمٌ	Merriage
D7	وَاللّٰهُ جَعَلَ لَكُمْ مِّنْ اَنْفُسِكُمْ اَزْوَاجًا وَّجَعَلَ لَكُمْ مِّنْ اَزْوَاجِكُمْ بَنِيْنَ وَحَفَدَةً وَّرَزَقَكُمْ مِّنَ الطَّيِّبٰتِ ۗ اَفَبِالْبَاطِلِ يُؤْمِنُوْنَ وَبِنِعْمَتِ اللّٰهِ هُمْ يَكْفُرُوْنَ	Merriage
D8	وَاللّٰهُ خَلَقَكُمْ مِّنْ تُرَابٍ ثُمَّ مِنْ نُّطْفَةٍ ثُمَّ جَعَلَكُمْ اَزْوَاجًا ۗ وَمَا تَحْمِلُ مِنْ اُنْثٰى وَلَا تَضَعُ اِلَّا بِعِلْمِهٖ ۗ وَمَا يُعَمَّرُ مِنْ مُّعَمَّرٍ وَّلَا يُنْقَصُ مِنْ عُمُرِهٖٓ اِلَّا فِيْ كِتٰبٍ ۗ اِنَّ ذٰلِكَ عَلَى اللّٰهِ يَسِيْرٌ	Merriage
D9	وَاسْتَعِيْنُوْا بِالصَّبْرِ وَالصَّلٰوةِ ۗ وَاِنَّهَا لَكَبِيْرَةٌ اِلَّا عَلَى الْخٰشِعِيْنَ	Sholat

Accredited "Rank 4"(Sinta 4), DIKTI, No. 36/E/KPT/2019, December 13th 2019.

Document	Verse	Class
D10	وَاَقِيْمُوا الصَّلٰوةَ وَاٰتُوا الزَّكٰوةَ ۗ وَمَا تُقَدِّمُوْا لِاَنْفُسِكُمْ مِّنْ خَيْرٍ تَجِدُوْهُ عِنْدَ اللّٰهِ ۗ اِنَّ اللّٰهَ بِمَا تَعْمَلُوْنَ بَصِيْرٌ	Sholat
D11	اَجَعَلْتُمْ سِقَايَةَ الْحَاۤجِّ وَعِمَارَةَ الْمَسْجِدِ الْحَرَامِ كَمَنْ اٰمَنَ بِاللّٰهِ وَالْيَوْمِ الْاٰخِرِ وَجٰهَدَ فِيْ سَبِيْلِ اللّٰهِ ۗ لَا يَسْتَوٗنَ عِنْدَ اللّٰهِ ۗ وَاللّٰهُ لَا يَهْدِى الْقَوْمَ الظّٰلِمِيْنَ	Hajj
D12	وَمِنْ كُلِّ شَيْءٍ خَلَقْنَا زَوْجَيْنِ لَعَلَّكُمْ تَذَكَّرُوْنَ	Merriage
D13	وَمَا خَلَقْتُ الْجِنَّ وَالْاِنْسَ اِلَّا لِيَعْبُدُوْنِ	Sholat
D14	وَجَاهِدُوْا فِى اللّٰهِ حَقَّ جِهَادِهٖ ۗ هُوَ اجْتَبٰىكُمْ وَمَا جَعَلَ عَلَيْكُمْ فِى الدِّيْنِ مِنْ حَرَجٍ ۗ مِلَّةَ اَبِيْكُمْ اِبْرٰهِيْمَ ۗ هُوَ سَمّٰىكُمُ الْمُسْلِمِيْنَ ۙ مِنْ قَبْلُ وَفِيْ هٰذَا لِيَكُوْنَ الرَّسُوْلُ شَهِيْدًا عَلَيْكُمْ وَتَكُوْنُوْا شُهَدَاۤءَ عَلَى النَّاسِ ۚ فَاَقِيْمُوا الصَّلٰوةَ وَاٰتُوا الزَّكٰوةَ وَاعْتَصِمُوْا بِاللّٰهِ ۗ هُوَ مَوْلٰىكُمْ ۚ فَنِعْمَ الْمَوْلٰى وَنِعْمَ النَّصِيْرُ	Sholat
D15	وَاٰتُوا النِّسَاۤءَ صَدُقٰتِهِنَّ نِحْلَةً ۗ فَاِنْ طِبْنَ لَكُمْ عَنْ شَيْءٍ مِّنْهُ نَفْسًا فَكُلُوْهُ هَنِيْٓـًٔا مَّرِيْٓـًٔا	Merriage
D16	وَاِنْ خِفْتُمْ اَلَّا تُقْسِطُوْا فِى الْيَتٰمٰى فَانْكِحُوْا مَا طَابَ لَكُمْ مِّنَ النِّسَاۤءِ مَثْنٰى وَثُلٰثَ وَرُبٰعَ ۚ فَاِنْ خِفْتُمْ اَلَّا تَعْدِلُوْا فَوَاحِدَةً اَوْ مَا مَلَكَتْ اَيْمَانُكُمْ ۗ ذٰلِكَ اَدْنٰٓى اَلَّا تَعُوْلُوْا ۗ	Merriage
D17	وَلَا تَنْكِحُوا الْمُشْرِكٰتِ حَتّٰى يُؤْمِنَّ ۗ وَلَاَمَةٌ مُّؤْمِنَةٌ خَيْرٌ مِّنْ مُّشْرِكَةٍ وَّلَوْ اَعْجَبَتْكُمْ ۚ وَلَا تُنْكِحُوا الْمُشْرِكِيْنَ حَتّٰى يُؤْمِنُوْا ۗ وَلَعَبْدٌ مُّؤْمِنٌ خَيْرٌ مِّنْ مُّشْرِكٍ وَّلَوْ اَعْجَبَكُمْ ۗ اُولٰٓئِكَ يَدْعُوْنَ اِلَى النَّارِ ۖ وَاللّٰهُ يَدْعُوْٓا اِلَى الْجَنَّةِ وَالْمَغْفِرَةِ بِاِذْنِهٖ ۚ وَيُبَيِّنُ اٰيٰتِهٖ لِلنَّاسِ لَعَلَّهُمْ يَتَذَكَّرُوْنَ	Merriage
D18	فَجَعَلَ مِنْهُ الزَّوْجَيْنِ الذَّكَرَ وَالْاُنْثٰى ۗ	Merriage
D19	وَمَآ اُمِرُوْٓا اِلَّا لِيَعْبُدُوا اللّٰهَ مُخْلِصِيْنَ لَهُ الدِّيْنَ ۙ حُنَفَاۤءَ وَيُقِيْمُوا الصَّلٰوةَ وَيُؤْتُوا الزَّكٰوةَ وَذٰلِكَ دِيْنُ الْقَيِّمَةِ ۗ	Sholat
D20	وَاَقِمِ الصَّلٰوةَ طَرَفَيِ النَّهَارِ وَزُلَفًا مِّنَ الَّيْلِ ۗ اِنَّ الْحَسَنٰتِ يُذْهِبْنَ السَّيِّاٰتِ ۗ ذٰلِكَ ذِكْرٰى لِلذّٰكِرِيْنَ	Sholat

The data in the table above are several documents before being classified and will then preprocess the documents.

3.2 Preprocessing Al-Quran Documents

The stages carried out in processing Al-Quran documents before classifying documents are Pre-processing where these stages aim to weight the terms which will then be continued with the document classification process while the stages are as follows:

a. The process of separating several words (tokenization)

Tokenization is a process of separating text from several rows of words in a sentence, paragraphs or in a page which later becomes a token or single word snippet (termed word). This stage also removes certain characters such as punctuation marks and then changes all tokens to small words (lower-cases).

b. Filtering Process,(Filtration)

Accredited "Rank 4" (Sinta 4), DIKTI, No. 36/E/KPT/2019, December 13th 2019.

Filtration is a process that is determined by the term to be used for the process document pre-presentation so that it can describe all the contents of the document and can distinguish one document from another in collection. Term conversion process to the basic form of the word (Stemming).

c. Term Weighting Process (Term Weighting)

Term Weighting is a process to give a weight according to weighting rules that have been selected, whether weighting locally, globally or a combination of both. In this case, many applications apply a combined weighting of the two, namely the multiplication of local weights or also called term and global inverse document frequency, abbreviated as tf-idf.

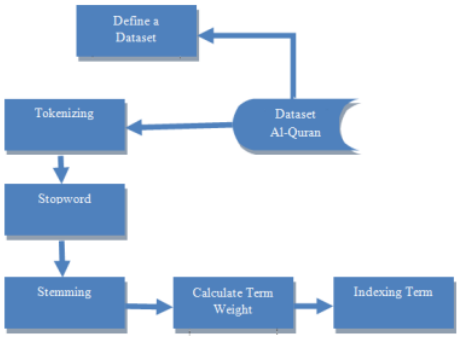

Fig 2: Preprocessing Document Text Chart

From some of the documents listed in table 2, the stages of preprocessing will be carried out as in Figure 2 after doing some screening and stemming, the terms and the appearance of TF will be obtained. After the process of calculating the inverse document frequency (IDF) as the formula for equation 2.2. from the results of the calculation, the next step is to classify paragraphs into several classes.

3.3 Document Classification

The next stage is the classification process after doing the text preprocessing stage. In this study, the process of determining the classification of Al-Quran documents uses Naïve Bayes. The stages before the classification process are:

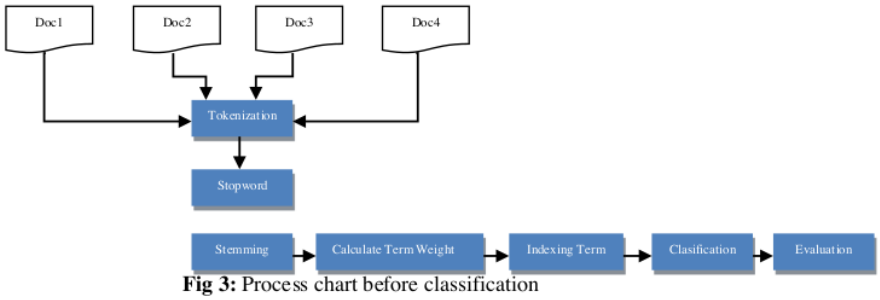

Fig 3: Process chart before classification

In the picture above, it is explained that the stages after pre-processing text will be continued with the classification stage, namely using Naïve Bayes which is then forwarded to the evaluation process. The formula for the Naïve Bayes algorithm is as follows:

$$P(Wk \mid Class) = \frac{nk + 1}{n + vocabulary}$$

Wk = Term Type
Class = Category Class
Nk = Number of occurrences of term in category
N = Number of term that appear in the category
Vocabulary = Sum of all terms

Accredited "Rank 4"(Sinta 4), DIKTI, No. 36/E/KPT/2019, December 13th 2019.

From the classification process with sample data 20 documents and 3 categories produce accuracy, recall and prescession as follows:

Table 3
Results from a sample of 20 documents with 3 categories

Category	Rercall	Precission	Accuracy
Sholat	90.00%	81.82%	70.00%
Hajj	62.50%	71.43%	
Merriage	0.00%	0.00%	

3.4 Evaluation and Results

In this stage, several series of trials and evaluations will be explained. That is the stages of training and testing of a number of Al-Quran documents are carried out as for the process as above and a test process will be carried out for other verses. This trial process will use data from 100 Al-Quran documents from a number of trials so that the accuracy, recall value, and precision can be obtained as follows:

Table 4
Result of the experiment

	True Marriage	True Sholat	True Hajj	Class Precission
Marriage Prediction	33	7	1	80.49%
Sholat Prediction	7	36	4	76.60%
Hajj Prediction	1	3	8	66.67%
Class Recall	80.49%	78.26%	61.54%	

From the table of prediction results, it can be concluded that

a. The value of True Positive (TP) for Marriage is 33, Sholat7, Hajj 1, and the value of other than TP is False Negative (FN).
b. While the recall class is a column that has a large value for the right classification, for example, the recall class from a marriage with the right classification is as follows:

$$TP = \frac{TP}{TP+FN} = \frac{33}{33+7+1} = \frac{33}{41} = 0.8048 = 80.49\%$$

c. While theclass precision is the row that contains the exact classification value. For example the class precision of the proper classification of marriage is as follows:

$$TP = \frac{TP}{TP+FN} = \frac{33}{33+7+1} = \frac{33}{41} = 0.8048 = 80.49\%$$

d. While accuracy is the result percentage from the correct classification that can be obtained by means.

$$TP = \frac{TP+FN}{TP+FN+FP+FN} = \frac{33+36+8}{33+7+1+7+36+3+1+4+8} = \frac{77}{100} = 0.77 = 77\%$$

Table 5
Text classification accuracy results

Category	Rercall	Precission	Accuracy
Sholat	80.49%	80.49%	77.00%
Hajj	78.26%	76.60%	
Merriage	61.54%	66.67%	

Accredited "Rank 4" (Sinta 4), DIKTI, No. 36/E/KPT/2019, December 13th 2019.

4. Conclusion

Based on the results of experiments that have been carried out, namely the classification of the Arabic text of the verses of Al-Qur'an can be explained that the Naïve Bayes Algorithm for classifying Arabic texts of Al-Quran verses produces a good accuracy that is 77.00%.

References

[1] Al-Qur'an Dan Terjemah Departemen Agama RI, Bandung: Jum'atul 'Ali-Art, 2005, p. 262.
[2] A. Nizhan, Buku Pintar Al-Qur'an, Jakarta Selatan: Qultum Media, 2008, p. 10.
[3] A. M. El-Halees, "Arabic Text Classification Using Maximum Entropy," The Islamic University Journal (Series of Natural Studies and Engineering), vol. 15, no. 1, pp. 157-167, 2007.
[4] A. M. El-Halees, "Mining Arabic Association Rules for Text Classification," 2006.
[5] A. Al-Zoghby, A. . S. Eldin, N. A. Ismail and T. Hamza, "Mining Arabic Text Using Soft Matching," Computer Engineering and Systems, pp. 421-426, 2007.
[6] A. Hamzah, "Klasifikasi teks dengan naïve bayes classifier (nbc) untuk pengelompokan teks berita dan abstract akademis," Prosiding Seminar Nasional Aplikasi Sains & Teknologi (SNAST) Periode III, no. 2011, pp. 269-277, 2012.
[7] A. S. Altheneyan and M. E. B. Menai, "Naïve Bayes classifiers for authorship attribution of Arabic texts," Journal of King Saud University - Computer and Information Sciences, vol. 26, no. 4, pp. 473-484, 2014.
[8] B. Ramesh and J. G. R. Sathiaseelan, "An Advanced Multi Class Instance Selection based Support Vector Machine for Text Classification," Procedia Computer Science, vol. 57, pp. 1124-1130, 2015.
[9] C. Darujati and A. . B. Gumelar, "Pemanfaatan Teknik Supervised Untuk Klasifikasi Teks Bahasa Indonesia," Link, vol. 16, no. 1, pp. 1-8, 2012.
[10] F. Harrag, E. El-Qawasmeh and P. Pichappan, "Improving Arabic Text Categorization using Decision Trees," p. 110 – 115, 2009.
[11] H. Faozi and E. El-Qawasmah, "Neural Network for Arabic text classification," 2009 Second International Conference on the Applications of Digital Information and Web Technologies,pp. 778-783, 2009.
[12] H. Sawaf, J. Zaplo and H. Ney, "Statistical Classification Methods for Arabic News Articles," Natural Language Processing in ACL2001, Toulouse, p. 6, 2001.
[13] L. Khreisat, "A machine learning approach for Arabic text classification using N-gram frequency statistics," Journal of Informetrics, vol. 3, no. 1, pp. 72-77, 2009.
[14] L. Khreisat, "A machine learning approach for Arabic text classification using N-gram," Journal of Informetrics, vol. 3, no. 1, pp. 72-77, 2009.
[15] M. El koudi, A. Bensaid and T.-e. Rachidi, "Automatic Arabic Document Categorization Based on the Naïve Bayes Algorithm," Proceedings of COLING 20th Workshop on Computational Approaches to Arabic Script-based Language, pp. 51-58, 2004.
[16] M. Bilal, H. Israr, M. Shahid and A. Khan, "Sentiment classification of Roman-Urdu opinions using Naïve Bayesian, Decision Tree and KNN classification techniques," Journal of King Saud University - Computer and Information Sciences, 2015.
[17] M. Q. Shihab, Membumikan Al-Quran, Bandung: Mizan, 1996, p. 1.
[18] R. Al-Shalabi, G. Kanaan and M. H. Gharaibeh, "Arabic text categorization using KNN," computer science and information, 2006.

Accredited "Rank 4"(Sinta 4), DIKTI, No. 36/E/KPT/2019, December 13th 2019.

20220611

ORIGINALITY REPORT

22%	19%	5%	17%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	site.iugaza.edu.ps Internet Source	9%
2	Submitted to Universitas Ibn Khaldun Student Paper	5%
3	iugspace.iugaza.edu.ps Internet Source	1%
4	quranlearning.info Internet Source	1%
5	www.iocscience.org Internet Source	1%
6	Submitted to The Scientific & Technological Research Council of Turkey (TUBITAK) Student Paper	1%
7	Submitted to Al-Nahrain University Student Paper	1%
8	Submitted to UIN Syarif Hidayatullah Jakarta Student Paper	1%
9	Submitted to State Islamic University of Alauddin Makassar	1%

Student Paper

10	Submitted to UIN Maulana Malik Ibrahim Malang Student Paper	1%
11	Submitted to UIN Sunan Gunung DJati Bandung Student Paper	1%

Exclude quotes	Off	Exclude matches	< 1%
Exclude bibliography	On		