

# Implementation of Text Mining Classification as a Model in the Conclusion of Tafsir Bil Ma'tsur and Bil RA'YI Contents

Afrizal Nur, Mustakim, Suja'i Syarifandi, Saidul Amin

**Abstract:** Studies related to Tafsir Qur'an have only been carried out based on a manual system or on application-based development. The purpose of this study is to build an application which can classify type of interpretation automatically into two classes, tafsir Bil Ma'tsur and tafsir Bil Ra'yi, can provide user convenience in the Al-Qur'an. KNN algorithm is a reliable algorithm in the classification process, also has many parts of the algorithm. This study was done by applying K-Nearest Neighbor (KNN) algorithm with an accuracy of 98.12%. However KNN had been compared firstly with Modified K-Nearest Neighbor (MKNN) and Fuzzy K-Nearest Neighbor (FKNN) algorithms, where the two algorithms had 98.01% and 88.3% accuracy respectively. MKNN is the best algorithm with the highest accuracy, but also has a high error value by 4.3% which is higher than KNN, 1.9%. From the research conducted, the more text documents used in KNN modeling, the higher accuracy will be. Therefore, in the implementation of this application KNN is used as modeling in the conclusion of Tafsir Al-Qur'an. Tests performed with BlackBox Testing and User Acceptance Test reached value of 100% and 98.8%.

**Index Terms:** Al-Qur'an, Classification, Fuzzy K-Nearest Neighbor, K-Nearest Neighbor, Tafsir.

## I. INTRODUCTION

The Qur'an is a guide book of hudan li an-nâs and rahmatan li al-'âlamîn. Al-Qur'an as hudan li an-nâs berarti bahwa Al-Qur'an means that the Qur'an is the way of life for Muslims. The purpose of the Qur'an as rahmatan li al-'âlamîn is a blessing for all nature, without exception Based on these two main functions; hudan and rahmatan, Al-Qur'an as the divine words revealed by Allah swt., to the Messenger of Allah, using one of the chosen languages, namely Arabic [1]. Because it is very unlikely that the two main goals can be realized properly except by using language that is easily understood by humans themselves. The goal is for humanity able to understand the contents of Qur'an, according to their respective capacities and abilities [2]. However, there are still many people, especially Muslims, who have not been able to understand the meanings and the

**Revised Manuscript Received on October 15, 2019.**

**Afrizal Nur**, Department of Science Al-Qur'an and Tafsir, Faculty of Ushuluddin, Universitas Islam Negeri Sultan Syarif Kasim Riau, Pekanbaru, Indonesia.

**Mustakim**, Department of Information System, Faculty of Science and Technology, Universitas Islam Negeri Sultan Syarif Kasim Riau, Pekanbaru, Indonesia.

**Suja'i Sarifandi**, Department of Science Al-Qur'an and Tafsir, Faculty of Ushuluddin, Universitas Islam Negeri Sultan Syarif Kasim Riau, Pekanbaru, Indonesia.

**Saidul Amin**, Department of Faith and Philosophy, Faculty of Ushuluddin, Universitas Islam Negeri Sultan Syarif Kasim Riau, Pekanbaru, Indonesia.

contents of Qur'an well, so that they have not been able to make Qur'an as hudan and rahmatan for themselves, families and the people. After the death of Rasullullah, the authority to understand the Qur'an was held by the Companions or Sahabah, Tabi'in, and Tabi'At-Tabi'in, until present scholars and Muslim scholars, of course with special conditions are required to understand and interpret Qur'an as the Book of Hidayah [2].

That approach was later known as the Science of Interpretation. The person who masters the knowledge and has the ability to interpret the Qur'an is called the mufassir [1]. In truth, the term of mufassir is just another name of mubayyin which was first known, because the meaning and understanding are the same. But at this time, the title of mufassir for interpreter was more popular than the term of mubayyin [1].

In the perspective of Koranic Sciences Study, at least two interpretive terminologies are often used, namely Tafsir Bi Al-Ma'tsur and Tafsir Bi Ar-Ra'yi. The interpretation of Bi Al-Ma'tsur is interpreted as an interpretation carried out by way of history, that is, interpretations originating from the Qur'an, Hadith, History of Sahabah r.a. and History of Tabi'in r.a. While the Bi Ar-Ra'yi Interpretation is defined as an attempt to uncover the contents of the Qur'an through ijthad which is done by appreciating the existence of reason [3]. The existence of mind which is also a representation of human knowledge is used as a basic reference in concluding the type of interpretation manually which then will be implemented into a media such as the internet or computer.

The development of computational technology can be used for various purposes of extracting information contained in a data. With the advent of technology such as Data Mining and Big Data, of course can help solve a problem both technically and in case studies Data Mining is a technique that applies algorithms to solve complex problems [4]. Data Mining techniques commonly used in problem solving include Classification and Clustering [5]. Among the two techniques, there is the K-Nearest Neighbor (KNN) Algorithm one of the Top Ten Algorithm [6].

Related to Data Mining, one popular trend in recent years is Text Mining. Text Mining is a group of documents in the form of irregular text which forms a data arranged by keywords. The existence of Text Mining can be used for various purposes such as getting information from Twitter, websites, electronic documents and some structured text documents such as al-quran, hadith and tafsir.



Text mining technique is able to extract information based on the structure of words and sentences, the relationship between one word and another word, as well as in the form of association relationships.

In the text structure of the archipelago tafsir there are two classes which are determined as targets in the process of classification, namely Tafsir Bil Ma'tsur and Tafsir Bil Ra'yi. The four types of word groupings used consist of the types of tafsir sources, namely (1) the Qur'an with the Qur'an, (2) the Qur'an with the Hadith, (3) the Qur'an with Sahabah's opinions and (4) The Qur'an with the opinion of Ta bi'in. The four types of Tafsir sources will be matched words in each row of tafsir documents to group and classify in the form of tracking the rules which have been determined by a certain weighting. Data mining technique used in this study to classify text on the tafsir of archipelago is K-Nearest Neighbor algorithm (KNN).

KNN is an algorithm that works by calculating the closest distance between data attributes [7], it has advantages in terms of high-performance computing [8], a simple algorithm and resilient to various characteristics of large data [9] also have good accuracy compared to other algorithms [8]. Several studies have been carried out by applying KNN for text mining, conducted by Yong, et al state that the classification of text using KNN has high accuracy with small algorithm complexity [10], then in 2016 KNN was used to classify the document text by modified KNN to K-Nearest Neighbor Improvement (IKNN), from the study stated that the accuracy of the IKNN was better than the KNN [11]. Based on its development, besides IKNN there are two other algorithms namely Modified K-Nearest Neighbor (MKNN) and Fuzzy K-Nearest Neighbor (FKNN). Where the two algorithms are part of the KNN development that is capable and reliable in completing supervised cases in Data Mining [12]. MKNN algorithm is an improvement in the performance and reliability of KNN method where each sample data will be validated to overcome the outlier data so that the weighting of each data can provide more reliable results [13]. Whereas FKNN is also a development of KNN combined with fuzzy theory so that a data has membership values in each class by considering the degree of membership in the interval (0,1) [14].

In this study, Tafsir Bil Ma'tsur and Tafsir Bil Ra'yi became a class in the classification of text Surah Al-Baqoroh in the tafsir of Al-Qur'an, Al-Misbah M Quraish Shihab. The class is used as a reference in concluding several new verses which have the same pattern based on the classification of Surah Al-Baqoroh. The main objective of this research is to design and build Data Mining Applications in concluding interpretation methodology using data from the best tafsir books. Besides implementing algorithms, this study also compares the KNN and its two derived algorithms, MKNN and FKNN. So that at the end of this study the best algorithm will be obtained to infer Tafsir Bil Ma'tsur and Tafsir Bil Ra'yi. Therefore, the main process of this study is a comparison of algorithms, applying algorithms with the best algorithmic results and implementation into the mobile application. In addition, the importance of this research is that, so far the implementation of the Qur'an in computers has been limited to the search, pronunciation and reading techniques, however, the absence of a special technique for grouping and classifying has never been done, especially in relation to determine Tafsir Bil Ma'tsur and Tafsir Bil Ra'yi.

## II. METHOD

The methodology in this study is divided into several main stages, preprocessing data, processing data, comparing the best algorithms and analyzing the best results. The main material in this study is Surah Al-Baqoroh's in Al-Misbah M Quraish Shihab's Tafsir of the Qur'an as data text that will be processed to produce an information. Likewise, the interpretation of Tafsir Bil Ma'tsur and Tafsir Bil Ra'yi became the main class in the classification process using KNN, MKNN and FKNN algorithms. Methodology can be shown in the following Figure 1.

### A. Tafsir Bi al Ma'tsur

Bi al-Ma'tsur's interpretation has two names. Sometimes it is called At-Tafsir bi al-Ma'tsur, and sometimes it is also called At-Tafsir bi an-Naqli. Al-Ma'tsur also means the hadith narrated by salaf scholars to khalaf scholars. So that it can be understood that al-Ma'tsur stands on the system of hadith narration [3].

The meaning of Tafsir bi al-Ma'tsur itself can be classified into two main meanings. First, everything mentioned in the Qur'an, both in the form of explanations and details of some verses in Qur'an. Second, everything narrated from the Messenger of Allah, and from the Sahabah r.a., even from tabi'in, in the form of explanations and information that are directly related to the verse of the Qur'an [15].

### B. Tafsir bi ar-Ra'yi

Tafsir bi ar-ra'yi is the tafsir of Qur'an verses through the process of ijtihaad which uses the role of reason, adheres to the rules of language and customs of Arabs. The interpretation of bi ar-ra'yi exists after the end of Salaf period around the 3rd century Hijriah and at that time Islamic civilization progressed and developed, so that various mahzab emerged among Muslims [15]. Each mahzab tries to convince people in order to develop their understanding.

Supported by the interpreters who have mastered various disciplines of science, then in the process of interpretation they tend to include the results of their own thoughts and discussions that are different from other interpreters. For example, Zamakhsyari in its interpretation of the verse of Qur'an, is tend into Balaghah. Unlike the Qurtubi, which is more focused on discussing aspects of sharia law [16].

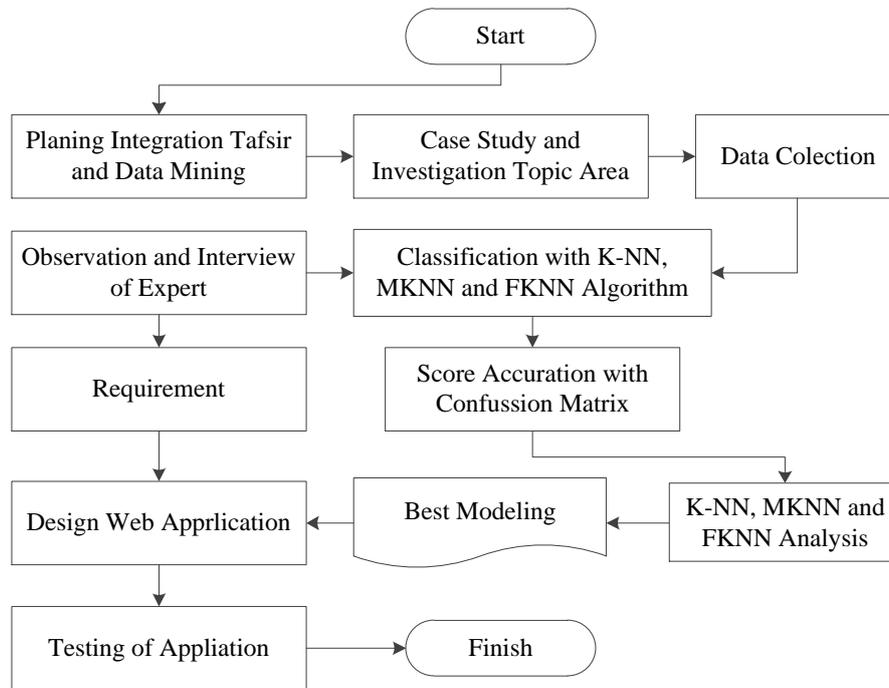


Figure 1. Research Methodology

### C. Text Mining

Text mining can be interpreted as the discovery of new and unknown information by computers, by automatically extracting information from different sources. The key to this process is combining information which was successfully extracted from various sources [17]. Whereas according to Harlian, text mining has a definition of mining data in the form of text where data sources are usually obtained from documents, and the purpose is to search for words that can represent the contents of the document so that it can analyze the connection between documents [18][19]. Text mining refers to the process of taking high-quality information from the text. High-quality information is usually obtained through forecasting patterns and trends through means such as learning statistical patterns [20].

### D. K-Nearest Neighbor (KNN)

This algorithm was first introduced by Fix and Hodges in 1951 and 1952 [21]. This algorithm is also one of the lazy learning techniques. KNN is done by finding groups of objects in the closest (similar) training data to objects on new data or testing data [22]. KNN is a data mining classification method based on learning by analogy. The training data sample has numerical dimension attributes. Each sample is a point in n-dimensional space. All training samples are stored in n-dimensional space. When testing data, it will search for the closest k value to the test data. Proximity is defined in terms of Euclidean distance between two points  $X=(x_1, x_2, \dots, x_n)$  dan  $Y=(y_1, y_2, \dots, y_n)$  [23][24].

### E. Modified K-Nearest Neighbor (MK-NN)

MKNN is a classification method based on proximity to training data. MKNN is an improvement in the performance and reliability of KNN method where each sample data will be validated to overcome the outlier data so that the weighting of each data can provide reliable results. In the MKNN method, each training data must be calculated as the validity value where the calculation process is based on the nearest neighbor [25]. Validity in each training data will be

used as a variable in carrying out the weighting in the next process. The MKNN algorithm is an enhancement algorithm from KNN Algorithm, a method for classifying objects based on learning data with the closest distance to the object. The steps of MKNN Algorithm plus Validity and Weight Voting calculations [26].

$$Validity(x) = \frac{1}{H} + \sum_{i=1}^H S(\text{lbl}(x), \text{lbl}(N_i(x))) \quad (1)$$

Note; H is the closest number of points,  $L_{bh}(x)$  is class x,  $N_i(x)$  is the class label nearest point x. Equation 1 is used to calculate the similarity of data x with the nearest neighbor data. The equation of the function S is defined in equation 2 below [26]:

$$S(a,b) = \begin{cases} 1, & a = b \\ 0, & a \neq b \end{cases} \quad (2)$$

In MKNN method, the first weight of each neighbor is calculated using  $1/(de + 0.5)$ . Then, the validity of each data in the training data is multiplied by the weight based on Euclidean distance. So that MKNN method will obtained the weight voting equation for each neighbor as in equation 3 below [26]:

$$W(x) = Validity(x) \times \frac{1}{(de + 0,5)} \quad (3)$$

### F. Fuzzy K-Nearest Neighbor (FK-NN)

FKNN is a classification method used to predict testing data using the value of membership in each class. Then the class is taken with the highest membership value from testing data as the prediction result class. FKNN is a classification method that combines Fuzz techniques with K-Nearest Neighbor classifier. The advantage is that vector membership values should provide a guaranteed level of classification results [14][27][28].

Fuzzy KNN is a combination of fuzzy methods and k-nearest neighbors in delivering class label to predicted data testing. The FKNN method can make predictions explicitly on data testing based on the nearest neighbor. In addition, it also predicts classes by giving membership values [29]. In fuzzy, a data has a membership value in each class that is different from the value of membership at the interval (0,1) [29]. This concept was introduced and first published by Lotfi A. Zadeh, a professor from the University of California at Berkeley in 1965. Fuzzy logic uses language expressions to describe variable values [14]. Fuzzy logic works by using the degree of membership value which is then used to determine the results to be produced based on predetermined specifications [30].

This FKNN algorithm plays an important role in removing ambiguity in classification. In addition, an instance will have a degree of membership in each class so that it will give more confidence to an instance in one class. Before calculating the membership value in FKNN, the process is first carried out using the following equation 4 [29][30]:

$$U_{ij} = \begin{cases} 0,51 + \left(\frac{n_j}{n}\right) * 0,49, & \text{jika } j = 1 \\ \left(\frac{n_j}{n}\right) * 0,49 & \text{jika } j \neq 1 \end{cases} \quad (4)$$

Note;  $n_j$  is the number of class  $j$  members in data training  $n$ ,  $n$  is the amount of training data used, and  $j$  is the data class. Next, calculate the membership value of each class with equation 5 below [28]:

$$U_{i(x)} = \frac{\sum_{j=1}^k u_{ij} (\|x-x_j\|^{-2/(n-1)})}{\sum_{j=1}^k (\|x-x_j\|^{-2/(n-1)})} \quad (5)$$

Note;  $U_{ij}$  is the fuzzy membership value in testing sample ( $x, x_j$ ),  $k$  is the nearest neighbor value,  $j$  is the variable of membership in data testing and  $m$  is the weight, with  $m > 1$ .

### III. RESULT AND ANALYSIS

In terms of interpretation, Ibnu 'Asyur uses various method, including the method of tahlili (analysis), naqdi (critical), istidlali (argumentative), maudhu'i (thematic) and maqashidi (objective). According to him, the combination of these methods can be revealed by the answers of Al-Qur'an about various life problems, as well as can be used as evidence that the verses of the Qur'an are in line with the development of science and technology and the progress of human civilization.

Specifically, Ibnu 'Asyur explained beforehand about the surah to be interpreted; starting from the meaning of the surah, the place to descend the surah, the number of verses in the surah, the reason it was revealed, the priority of the surah, to the content of the surah in general. Then he combined a number of verses with the same theme, then interpret it using correlation analysis between verses or surah, linguistic analysis, related histories, and opinions of previous scholars. After summarizing, he then took the steps of ijihad using the method of istidlali, naqdi and maqashidi. This is the main source of knowledge in this study in concluding the Qur'an, especially by studying a number of structures. However, as related to the process of interpretation, the main role of an interpreter in this study is as a validator related to the conclusion of the verse in Qur'an. The main source used is the text of the Indonesian translation of each verse. So there may be several errors in interpretation, then it will be measured based on their accuracy.

Research related to text mining, integrated with the Science of Interpretation has several models and methods carried out both textually and technically. Research involving several interpreters from UIN Sultan Syarif Kasim Riau and researchers from Bandung Institute of Technology (ITB) has its own attraction, namely in the process of documenting the main sources based on the results of previous research, Surat Al-Baqoroh verses 1-120 [31]. This research starts from data collection, namely the translation of the letter Al-Baqoroh, simplification of text and preprocessing and comparing the three algorithms used.

Table 1. The Data Set Al-Qur'an Al-Baqoroh 266-269

No	Verse	Text	Type of Tafsir
1	266	Apakah ada salah seorang di antaramu yang ingin mempunyai kebun kurma dan anggur yang mengalir di bawahnya sungai-sungai; dia mempunyai dalam kebun itu segala macam buah-buahan, Kemudian datanglah masa tua pada orang itu sedang dia mempunyai keturunan yang masih kecil-kecil. Maka kebun itu ditiup angin keras yang mengandung api, lalu terbakarlah. Demikianlah Allah menerangkan ayat-ayat-Nya kepada kamu supaya kamu memikirkannya.	bi Ra'yi
2	267	Hai orang-orang yang beriman, nafkahkanlah (di jalan Allah) sebagian dari hasil usahamu yang baik-baik dan sebagian dari apa yang kami keluarkan dari bumi untuk kamu. dan janganlah kamu memilih yang buruk-buruk lalu kamu menafkahkan daripadanya, padahal kamu sendiri tidak mau mengambilnya melainkan dengan memincingkan mata terhadapnya. dan Ketahuilah, bahwa Allah Maha Kaya lagi Maha Terpuji	bi Ma'Tsur
3	268	Syaitan menjanjikan (menakut-nakuti) kamu dengan kemiskinan dan menyuruh kamu berbuat kejahatan (kikir); sedang Allah menjadikan untukmu ampunan daripada-Nya dan karunia. dan Allah Maha luas (karunia-Nya) lagi Maha Mengatahui	bi Ra'yi
4	269	Allah menganugerahkan Al hikmah (kefahaman yang dalam tentang Al Quran dan As Sunnah) kepada siapa yang dikehendaki-Nya. dan barangsiapa yang dianugerahi hikmah, ia benar-benar Telah dianugerahi karunia yang banyak. dan Hanya orang-orang yang berakallah yang dapat mengambil pelajaran (dari firman Allah).	bi Ra'yi

The cleaning process is done by changing the text to lowercase (casefold), removing all non-letter characters and deleting punctuation. From this stage, all data can be used, namely text in documents 1-286. This means that in this process there is no single document selected and in accordance with the original text. While the pre-processing is a process for preparing raw data before data analysis is

carried out. In general, preprocessing data is done by eliminating inappropriate data or converting data into a form that is easier to understand by the system. Some of the processes carried out on preprocessing are Tokenizing, Filtering, Stemming, and Tagging.



So as to get a word like the example of Al-Baqoroh verse 266 as follows “apa ada salah orang antara ingin punya kebun kurma anggur alir bawah sungai dia dalam segala macam buah kemudian datang masa tua sedang turun masih kecil maka tiup angin keras kandung api lalu bakar allah terang ayat kamu supaya pikir”.

For the next process, Term Frequency - Inverse Document Frequency and Vector Space Model are carried out so that the final data is obtained as follows:

**Table 2. Preprocessing Final Results**

No	Cosine	Verse	Type of Tafsir
1	0,7652	001	1
2	...	...	...
3	0,6288	266	1
4	0,5032	267	2
5	0,2921	268	1
6	0,7436	269	1
7	...	...	...
8	0,3535	286	1

**Analysis of KNN, MKNN and FKNN**

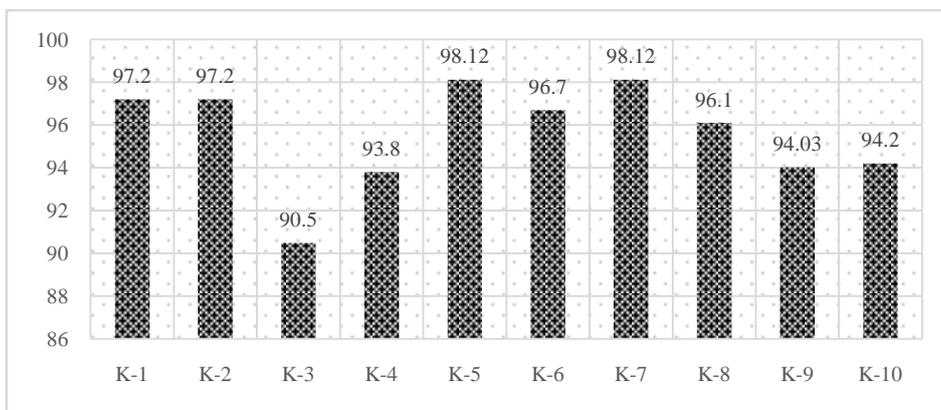
The classification of KNN, MKNN and FKNN algorithms were done by using 10 experiments on cross validation. The highest accuracy results based on matrix confusion became a recommendation to develop the application. The results of the classification of Al-Qur'an Al-Baqoroh verses 1-120 using K-NN algorithm stated that the study has the highest accuracy of 97.7% at k = 8 with a minimum error value of 1.9% [26]. But in this study the number of documents of 286 verses were able to increase the accuracy by 0.31% so that the best accuracy in KNN reached

98.01%. The increase in accuracy in each experiment using the KNN algorithm was very significant, except for the values k = 3 and k = 4, there was a decrease in accuracy. Therefore, it can be concluded that the larger the text data or text documents used can increase the accuracy of KNN. In comparison the accuracy of each k experiment on KNN can be shown in the following table 3:

**Table 3. Comparison of KNN Accuracy with 120 documents and 286 documents**

Fold	KNN with 120 documents (Nur dkk, 2018)	KNN with 286 documents
K-1	93.8	93.6
K-2	92.6	93.4
K-3	92.8	88.6
K-4	92.8	89.9
K-5	96.4	96.8
K-6	95.3	96.8
K-7	95.9	97.3
K-8	<b>97.7</b>	<b>98.0</b>
K-9	96.0	97.7
K-10	96.2	96.4

The results of MKNN classifications which are also based on 10 experiments, produced the highest accuracy of 98.12% but have a considerable error value of 4.3%. The highest accuracy is at k = 5 and k = 7. In MKNN algorithm, the determinant value in each iteration has a very small probability so that the resulting error may be greater. This has become the main problem in this algorithm in every trial process. Complete results of MKNN classification can be shown in the following picture 2:



**Figure 2. Accuracy in the MKNN Algorithm**

In contrast to KNN and MKNN, FKNN has a unique pattern in the case of text classification in Al-Qur'an which has the same accuracy value of more than 50% of the total experiments below the highest accuracy of KNN and MKNN, by 88.3% at k = 3 k = 5, k = 8 and k = 9. While the lowest accuracy value is not far from the highest value of 86.7%. FKNN's accuracy in detail can be shown in the following 3 image.

Of the three algorithms applied in this study, it was found that MKNN is an algorithm with the best accuracy, but this

research is only limited to conducting accuracy tests. The findings of this study indicate that in concluding the verses in the Qur'an in Surah Al-Baqoroh can use KNN or MKNN algorithm. But what needs to be considered is the error of each experiment is also became its own parameter. MKNN is the algorithm with the highest accuracy, but also has a high error value. While KNN also has high accuracy too, but the resulting error value is lower than MKNN. In detail, the comparison of accuracy from the three algorithms can be shown in the following table 4.

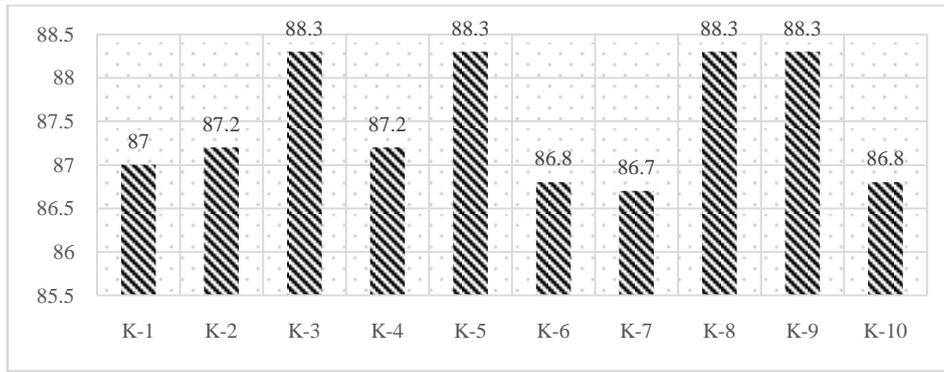


Figure 3. Accuracy in FKNN Algorithm

Table 4. Acuration of Comparison All Algorithm

K-Fold	KNN Old Research	KNN	MKNN	FKNN
K-1	92.8	93.6	97.2	87.0
K-2	92.6	93.4	97.2	87.2
K-3	92.8	88.6	90.5	88.3
K-4	92.8	89.9	93.8	87.2
K-5	96.4	96.8	98.1	88.3
K-6	95.3	96.8	96.7	86.8
K-7	95.9	97.3	98.1	86.7
K-8	97.7	98.0	96.1	88.3
K-9	96.0	98.7	94.03	88.3
K-10	96.2	96.4	94.2	86.8

Whereas if considered from the highest k value of the three algorithms KNN, MKNN and FKNN can be shown in the following image 4:

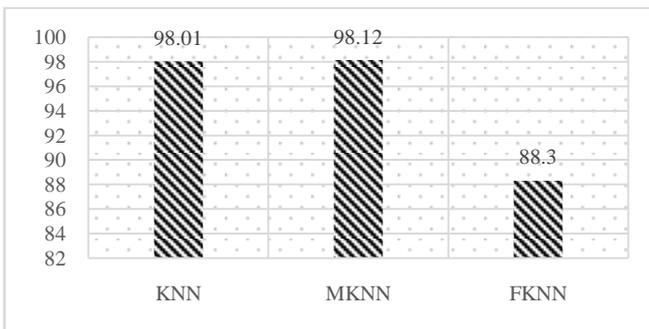


Figure 4. The highest accuracy in the KNN, MKNN and FKNN algorithms

If examined more deeply, from this comparison MKNN is a chosen algorithm which will be applied to mobile-based applications in the conclusion of Al-Qur'an by accuracy of 97.9%. However, it can be said that this algorithm cannot be applied globally or multi-case, because experiments are carried out only by using small data with a variety of text structures. Therefore to apply with different cases, MKNN cannot be the main reference to be applied in text mining because it has greater complexity. Besides that, the error value becomes a another obstacle in implementing MKNN algorithm, especially if the larger documents used allow a large number of errors to be generated.

**Application Analysis and Implementation**

The analysis starts from user requirements for students, lecturers and experts in the field of information technology. The analysis is divided into 3 main parts, namely based on Context Diagrams (CD), Data Flow Diagrams (DFD) and Entity Relationship Diagram (ERD). Context Diagram or CD is a description of the whole actor involved in the system, both input and output. In this case there are 2 actors, Admin,

who are in charge of inputting data on the system as a whole, and Users who have the task as tracer or information seekers on this system.

DFD is used to determine how to design the interface design and menu structure in application. In this research, system design used consists of several main points, they are: (1) 2 main displays, namely the interface for user in general without having to enter the system using a username and password, the user can only search for information related to the Qur'an comment. Unlike the second display, the user is required to enter by using a username and password. (2) There are 2 parts of search view, first, search in general, the user only input the keywords according to what is searched for, for example related to words or sentences. While for more detailed users can use the Detailed Search menu consisting of Letters and Verses, Types of Interpretations and Narrator. (3) In the design displayed a user can input searches based on detailed search without using keywords, meaning that the user only selects several dropdown menus available.

In addition, the testing stage is done by using blackbox testing techniques and User Acceptance Test (UAT) with each test value is 100%. All processes in the system are running well and the assessment is carried out by 20 UAT testers with system capabilities in interaction between system and users amounting is 98.8%, which means that only a few test respondents did not succeed.

**III. CONCLUSION**

From the discussion and analysis in this study, it can be concluded into two main parts, first, the comparison of KNN algorithms between previous research and this study has an increased of accuracy value by 0.31%, meaning that the more verses documents are used, the better the accuracy of KNN algorithm. Furthermore, the comparison between KNN, MKNN and FKNN has a high accuracy value in MKNN algorithm, which is equal to 98.12, higher than that of KNN and FKNN, 98.01% and 88.3%. But MKNN algorithm has a greater error than KNN which is 4.3% higher than KNN which is only 1.9%. Therefore, the implementation used in concluding the verses is KNN algorithm, the difference in accuracy between the best algorithms of MKNN is only 0.11% and the error ratio is better for KNN. Second, the application of KNN algorithm is applied using mobile application platform with value of blackbox testing is 100% and UAT testing is 98.8%, which means that this application is easy to use.



The most important thing of this application is being able to provide information related to the type of Tafsir, Tafsir Bil Ma'tsur and Tafsir Bil Ra'yi, applications that are built with web-based administrators are also able to predict other verses into these two classes, making it easier for users study interpretations correctly.

## REFERENCES

1. A. Nur, M. Lubis, and H. Ishak, "Sumbangan Tafsir al-Tahrir wa al-Tanwir Ibn 'Ashur dan Relasinya dengan Tafsir al-Mishbah M. Quraysh Shihab," AL-TURATH J. AL-QURAN AL-SUNNAH, vol. 2, no. 2, pp. 67–79, 2017.
2. A. Nur and M. Yasir, "THE INTERPRETATION OF VERSES 39 OF AN-NUR IN THE PERSPECTIVE OF TAFSIR AND SCIENCE."
3. A. Nur, "Konsep Wasathiyah Dalam Al-Quran;(Studi Komparatif Antara Tafsir Al-Tahrir Wa At-Tanwir Dan Aisar At-Tafasir)," J. An-Nur, vol. 4, no. 2, 2016.
4. D. Tomar and S. Agarwal, "A survey on Data Mining approaches for Healthcare," Int. J. Bio-Science Bio-Technology, vol. 5, no. 5, pp. 241–266, 2013.
5. S. G. Jacob and R. G. Ramani, "Evolving efficient clustering and classification patterns in lymphography data through data mining techniques," Int. J. Soft Comput., vol. 3, no. 3, p. 119, 2012.
6. R. Agrawal, "K-Nearest neighbor for uncertain data," Int. J. Comput. Appl., vol. 105, no. 11, 2014.
7. A. B. Hassanat, M. A. Abbad, G. A. Altarawneh, and A. A. Alhasanat, "Solving the problem of the K parameter in the KNN classifier using an ensemble learning approach," arXiv Prepr. arXiv1409.0919, 2014.
8. H. S. Khamis, K. W. Cheruiyot, and S. Kimani, "Application of k-nearest neighbour classification in medical data mining," Int. J. Inf. Commun. Technol. Res., vol. 4, no. 4, 2014.
9. S. B. Imandoust and M. Bolandraftar, "Application of k-nearest neighbor (knn) approach for predicting economic events: Theoretical background," Int. J. Eng. Res. Appl., vol. 3, no. 5, pp. 605–610, 2013.
10. Z. Yong, L. Youwen, and X. Shixiong, "An improved KNN text classification algorithm based on clustering," J. Comput., vol. 4, no. 3, pp. 230–237, 2009.
11. S. Gupta and R. Rani, "Improvement in KNN Classifier (imp-KNN) for Text Categorization," IJARCSSE, 2016.
12. M. Pangestu, B. Hidayat, and F. Oscandar, "Identifikasi Pola Rugae Palatina Untuk Klasifikasi Jenis Kelamin Manusia Dengan Citra Digital Menggunakan Metode Gabor Wavelet Dan Fuzzy K-nn," eProceedings Eng., vol. 3, no. 2, 2016.
13. M. Bannayan and G. Hoogenboom, "Weather analogue: a tool for real-time prediction of daily weather data realizations based on a modified k-nearest neighbor approach," Environ. Model. Softw., vol. 23, no. 6, pp. 703–713, 2008.
14. S. D. Nugraha, R. R. M. Putri, and R. C. Wihandika, "Penerapan Fuzzy K-Nearest Neighbor (FK-NN) Dalam Menentukan Status Gizi Balita," J. Pengemb. Teknol. Inf. dan Ilmu Komput. e-ISSN, vol. 2548, p. 964X.
15. [A. Nur, "Khazanah Dan Kewibawaan Tafsir Bi Al-Ma'tsur." Asa Riau, 2015.
16. M. S. Razali, "RASIONALITAS ALQURAN DALAM TAFSIR AN-NUR: STUDI PENAFSIRAN SURAH AL-NISA' [4]: 1," Mutawatir J. Keilmuan Tafsir Hadith, vol. 6, no. 1, pp. 179–205, 2016.
17. M. Hearst, "WordSeer: A Text Analysis Environment for Literature Study," 2016.
18. M. Harlian, "Text Mining," Diakses pada tanggal, vol. 27, 2006.
19. M. Harlian, "Machine Learning Text Kategorization," Austin Univ. Texas, 2006.
20. M. W. Berry and J. Kogan, Text mining: applications and theory. John Wiley & Sons, 2010.
21. O. Kramer, "K-nearest neighbors," in Dimensionality reduction with unsupervised nearest neighbors, Springer, 2013, pp. 13–23.
22. H. Parvin, H. Alizadeh, and B. Minati, "A modification on k-nearest neighbor classifier," Glob. J. Comput. Sci. Technol., 2010.
23. S. A. Dudani, "The distance-weighted k-nearest-neighbor rule," IEEE Trans. Syst. Man. Cybern., no. 4, pp. 325–327, 1976.
24. M. Mustakim and G. Oktaviani, "Algoritma K-Nearest Neighbor Classification Sebagai Sistem Prediksi Predikat Prestasi Mahasiswa," J. Sains dan Teknol. Ind., vol. 13, no. 2, pp. 195–202, 2016.
25. I. Gazalba and N. G. I. Reza, "Comparative analysis of k-nearest neighbor and modified k-nearest neighbor algorithm for data classification," in Information Technology, Information Systems and Electrical Engineering (ICITISEE), 2017 2nd International conferences on, 2017, pp. 294–298.
26. H. Parvin, H. Alizadeh, and B. Minaei-Bidgoli, "MKNN: Modified k-nearest neighbor," in Proceedings of the World Congress on Engineering and Computer Science, 2008, vol. 1.
27. K.-C. Chou and H.-B. Shen, "Predicting eukaryotic protein subcellular location by fusing optimized evidence-theoretic K-nearest neighbor classifiers," J. Proteome Res., vol. 5, no. 8, pp. 1888–1897, 2006.
28. H.-L. Chen, C.-C. Huang, X.-G. Yu, X. Xu, X. Sun, G. Wang, and S.-J. Wang, "An efficient diagnosis system for detection of Parkinson's disease using fuzzy k-nearest neighbor approach," Expert Syst. Appl., vol. 40, no. 1, pp. 263–271, 2013.
29. B. F. Billyan, A. Bhawiyuga, and R. Primananda, "Implementasi Metode Klasifikasi Fuzzy K-Nearest Neighbor ( FK-NN ) Untuk Fingerprint Access Point Pada Indoor Positioning," J. Pengemb. Teknol. Inf. dan Ilmu Komput., vol. 1, no. 11, pp. 1195–1205, 2017.
30. J. M. Keller, M. R. Gray, and J. A. Givens, "A fuzzy k-nearest neighbor algorithm," IEEE Trans. Syst. Man. Cybern., no. 4, pp. 580–585, 1985.
31. A. Nur, M. Mustakim, and C. Bella, "Application of Data Mining As Model In Concluding The Substance of Tafsir Bil Ma ' tsur and Bil Ra ' yi In Al-Tahrir Wa Al- Tanwir of Ibnu ' Asyur."

## AUTHORS PROFILE



**Afrizal Nur**, is a lecturer at Alquran and Tafsir Department at Sultan Syarif Qasim Riau State Islamic University. He is a subject expert at Tafsir discipline. Most of his research and publication are related with Tafsir especially an integrated study of Tafsir with data mining, expert system and bayesian network. He is also active as caretaker of a scientific consortium : Indonesian Alquran and Tafsir association (AIAT) 2017-2020.



**Mustakim**, is a lecturer at Departement of Information System Faculty of Science and Technology Universitas Islam Negeri Sultan Syarif Kasim Riau. Subject expert at Data Mining and Big Data. Most of his research and publication are integrated study of Data Mining with Tafsir, Data Mining with Education and Data Mining with Software Engineering.



**Suja'i Saripandi**, is a lecturer at Hadist department at Sultan Syarif Qasim Riau State Islamic University. He is a subject expert at Hadist discipline. His published research mostly related to Hadist discipline. He served as the Chief of Quality Assurance Committee of Ushuluddin Faculty in 2017.



**Saidul Amin**, is a lecturer at Faith and Philosophy department at Sultan Syarif Qasim Riau State Islamic University. He is a subject expert at Islamic philosophy. His published research mostly related to Islamic philosophy..