A Topical Classification of Quranic Arabic Text

Conference Paper · December 2013

4 authors, including:

Mohammed N. Al-Kabi
Zarqa University
115 PUBLICATIONS 528 CITATIONS

Belal Mustafa Abuata
Yarmouk University
35 PUBLICATIONS 40 CITATIONS

Izzat Alsmadi
Yarmouk University
223 PUBLICATIONS 511 CITATIONS

Some of the authors of this publication are also working on these related projects:

Web Technology View project

The future of programmable security controls View project

All content following this page was uploaded by Mohammed N. Al-Kabi on 13 June 2014.

The user has requested enhancement of the downloaded file. All in-text references underlined in blue are added to the original document and are linked to publications on ResearchGate, letting you access and read them immediately.
A Topical Classification of Quranic Arabic Text

Mohammed N. Al-Kabi
Faculty of Sciences & IT
Zarqa University
Zarqa, Jordan
malkabi@zu.edu.jo

Belal M. Abu Ata
CIS Department
IT & CS Faculty
Yarmouk University
Irbid, Jordan
belalabuata@yu.edu.jo

Heider A. Wahsheh
Computer Science Department
College of Computer Science
King Khalid University
Abha, Saudi Arabia
heiderwahsheh@yahoo.com

Izzat M. Alsmadi
Information Systems Department
College of Computer & Information Sciences
Prince Sultan University
Riyadh 11586, P. O. Box 66833, Saudi Arabia
ialsmadi@cis.psu.edu.sa

Abstract—The automatic information classification is an important tool used today in many aspects of our life. This tool is used in document classification, speech recognition, handwriting recognition, search engines, data mining, question-answering systems, etc. There are many conducted studies in the area of English and Arabic textual document/sentence classification. However, the literature has one or two primitive studies to classify the Holy Quran Ayāts (verses). This study aims to evaluate the effectiveness of four well-known classification algorithms (Decision Tree, K-Nearest Neighbor (K-NN), Support Vector Machine (SVM) and Naïve Bayes (NB)) to classify different Quranic Ayāts according to their topics.

Keywords- Data mining; Feature selection; SVM; Arabic text classification; Quran text classification; Machine learning;

I. INTRODUCTION

Islam as a religion is based mainly on two sources; first source is the holy Quran which represents the verbatim word of God (Allah) and final revelation to humanity. Therefore it is considered a highly esteemed literary source. The second source Islam based on is Sunnah (Prophet Muhammad sayings). Sunnah includes interpretations of the holy Quran, teachings and way of life of the prophet.

Holy Quran is considered the main miracle of Prophet Muhammad. Prophet Moses miracles includes parting of the Red sea, Prophet Jesus miracles includes curing patients and resurrection of the dead. Moses' miracles and Jesus' miracles are witnessed by some people who live at their time and it is a proof of those Witnesses, and holy Quran asserts the occurrence of these miracles. Holy Quran as a miracle of Prophet Muhammad is different from the miracles of his predecessors (Prophet Moses and Prophet Jesus), since it defies the Arabs in their language at all times. So it is an eternal miracle since it characterizes by it linguistic perfection and inimitability, true prophecies, and validation of recent scientific discoveries.

The literal meaning of the Arabic word "Quran" or "Qur an" or "(Al-Qur ān)" is "the recitation". The holy Quran revelation took over 23 years, where some parts of it revealed at Mecca city and other parts at Medina city in Saudi Arabia. Holy Quran consists of 114 Suras (chapters). Each Sura consists of a number of ayāt (verses). These Suras (Chapters) is varies widely in their lengths, where Sura lengths measured in the number of ayāt. The first sura (Al-Fatihā) which consists of seven verses only, while the second sura (Al-Baqarah) consists of 286 verses [1-3]. The total number of Quranic ayāt (verses) is 6,236 [4]. There are 77,430 Arabic words used in holy Quran, and the total number of unique Arabic words is 18,994. Also there are 12,183 stems; 3,382 lemmas and 1,685 roots [5].

Arabic language is used mainly in the Middle East and North Africa by around 300 million people these days. Arabic language is the 6th most spoken language worldwide, and it is one of united nation six official languages. It is number one Semitic language used today followed by Amharic language which is the official language of Federal Democratic Republic of Ethiopia, followed by Hebrew. Arabic language is really rich and deep, and this makes a number of Muslim scholars believe that this reason behind the revelation of original Quran was in Arabic. You can find many Arabic terms which differ in their meanings but their English translation is a single English word. Examples of such Arabic words are "الْقَلْب" and "الفَوَاد" which means heart in English, but in Arabic they have two different meanings. Also The Arabic words "الْإنسان" "الْفَلَس" "الْفَلْس" "الْفَلْس" have different meanings in Arabic and they are not synonyms, but the translator has to translate them to human being(s) in English. In Quran you can find palindrome like in Mudathir Sura (chapter) 74, the third āyw (verse) states (Declare the greatness only of your Lord, "لَرَيْكَ فَكْرَنَ"). The palindrome should be applied on non-vocalized Arabic text, in other words palindrome should be applied on Arabic text free from diacritical marks.

Holy Quran is originally written in Arabic without using diacritical marks. Arabic orthography was under development at the time of Qur'anic revelation. The evolution of Arabic alphabets started at the 7th century to end up in 8th century by Al-Khalil ibn Ahmad al-Farahidi. Therefore Arabic writings at the time of Quran revelation were un-dotted and free from hamzas and diacritical marks.

The literature lacks to studies about the automatic topical classification of Quranic Arabic ayāt (verses). This may be
due to difficulty of the Arabic language (Classical and Modern Standard Arabic), and the Quranic Arabic language. The holy Quran uses a spelling that is different from the Arabic spelling used these days in newspapers, schools, university, Arabic literature which is known as Modern Standard Arabic (MSA). Let us consider the following four Arabic words extracted from Al-Fatiha sura (Chapter 1) as shown in http://www.quranexplorer.com/quran/ and http://corpus.quran.com. The Dagger Alif or superscript Alif (дер Альф, " cavalier"
which is written as a short vertical stroke on top of an Arabic letter are used in Arabic words presented in table I. Table I exhibits clearly the difference in the spelling of holy Quran and MSA spelling for the same words.

<table>
<thead>
<tr>
<th>Sura (Chapter)</th>
<th>Ayah (Verse)</th>
<th>English meaning</th>
<th>Modern Standard Arabic (MSA)</th>
<th>Quranic Arabic</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>The Worlds</td>
<td>العالمين مالك</td>
<td>مالك العالمين</td>
</tr>
<tr>
<td>1</td>
<td>4</td>
<td>Owner</td>
<td>مالك</td>
<td>مالك</td>
</tr>
<tr>
<td>1</td>
<td>6</td>
<td>The path</td>
<td>الصراط خنجرة</td>
<td>الصراط</td>
</tr>
<tr>
<td>1</td>
<td>7</td>
<td>Path</td>
<td>الصراط خنجرة</td>
<td>الصراط</td>
</tr>
</tbody>
</table>

We fail to find a dataset for holy Quran that we can use in our studies. Therefore this and future studies will include building such dataset. This study is based on a sample of textual dataset of holy Quran. This textual dataset has Arabic Quranic ÿåt (verses) which identified by Muslim scholars as belonging to one of the three selected topics (classes) (Ignorant of religion, " توحيد") (Oneness of God, "الحاجزين"), (Penalty of Apostates, " جزاء المنافقين") and (Penalty of Apostates, " الجزاء المنافقين") in the spellings of the four Arabic words extracted from Al-Fatiha sura (Chapter 1).

<table>
<thead>
<tr>
<th>Sura (Chapter)</th>
<th>Ayah (Verse)</th>
<th>English meaning</th>
<th>Modern Standard Arabic (MSA)</th>
<th>Quranic Arabic</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>The Worlds</td>
<td>العالمين مالك</td>
<td>مالك العالمين</td>
</tr>
<tr>
<td>1</td>
<td>4</td>
<td>Owner</td>
<td>مالك</td>
<td>مالك</td>
</tr>
<tr>
<td>1</td>
<td>6</td>
<td>The path</td>
<td>الصراط خنجرة</td>
<td>الصراط</td>
</tr>
<tr>
<td>1</td>
<td>7</td>
<td>Path</td>
<td>الصراط خنجرة</td>
<td>الصراط</td>
</tr>
</tbody>
</table>

The explorer in the world of Quranic Web sites will notice different Arabic spellings which do not change the Arabic word, therefore the meaning and pronunciation is not change. Websites like http://tanzil.net/#1:1, and http://www.holyquran.net/cgi-bin/prepare.pl?ch=1 use different spellings of the four Arabic words extracted from Al-Fatiha sura (Chapter 1) as shown in table II. Those Websites use normal (Alif, "ال") instead of (Dagger Alif, " ألف خنجرة") in the spellings of the four Arabic words extracted from Al-Fatiha sura (Chapter 1).

The rest of the paper organized as follows: Section 2 presents a summary to a number of studies related to Arabic textual classification and in particular to the classification of the verses of holy Quran ÿåt (verses). Section 3 presents the adopted approach. Section 4 will show the results of testing the four classification algorithms under consideration. Section 5 presents our conclusions and future plans to improve this study.

<table>
<thead>
<tr>
<th>Sura (Chapter)</th>
<th>Ayah (Verse)</th>
<th>English meaning</th>
<th>Modern Standard Arabic (MSA)</th>
<th>Quranic Arabic</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>The Worlds</td>
<td>العالمين مالك</td>
<td>مالك العالمين</td>
</tr>
<tr>
<td>1</td>
<td>4</td>
<td>Owner</td>
<td>مالك</td>
<td>مالك</td>
</tr>
<tr>
<td>1</td>
<td>6</td>
<td>The path</td>
<td>الصراط خنجرة</td>
<td>الصراط</td>
</tr>
<tr>
<td>1</td>
<td>7</td>
<td>Path</td>
<td>الصراط خنجرة</td>
<td>الصراط</td>
</tr>
</tbody>
</table>

In table II the last two columns present fully spelling matched Arabic words. The last column in table II presents a fully vocalized (vowelised or vowelled) Arabic words, while the fourth column presents non-vocalized Arabic Words as used in MSA.

Although of the challenges of the Arabic language, there are a decent number of studies about the Arabic textual classification. Most of these studies used Modern Standard Arabic (MSA) datasets, such as the studies of [6-11] etc. We found only one primitive study in the literature dedicated to the topical classification of only two Suras (chapters) of the holy Quran [12].

We fail to find a dataset for holy Quran that we can use in our studies. Therefore this and future studies will include building such dataset. This study is based on a sample of textual dataset of holy Quran. This textual dataset has Arabic Quranic ÿåt (verses) which identified by Muslim scholars as belonging to one of the three selected topics (classes) (Ignorant of religion, " توحيد") (Oneness of God, "الحاجزين"), (Penalty of Apostates, " جزاء المنافقين") and (Penalty of Apostates, " الجزاء المنافقين") in the spellings of the four Arabic words extracted from Al-Fatiha sura (Chapter 1).

The rest of the paper organized as follows: Section 2 presents a summary to a number of studies related to Arabic textual classification and in particular to the classification of the verses of holy Quran ÿåt (verses). Section 3 presents the adopted approach. Section 4 will show the results of testing the four classification algorithms under consideration. Section 5 presents our conclusions and future plans to improve this study.

II. RELATED WORK

This section briefly surveys the work on Arabic textual classification. Also we present at that end of this section a summary to only one study conducted on the field of topical classification of holy Quran verses.

Duwairi [8] study is based on a dataset consisting of 1,000 Arabic documents which vary in length and writing style. Those documents are distributed equally into 10 predefined classes. In her study a distance-based classifier is tested using the 1,000 Arabic documents. Arabic stemming is adopted in Duwairi [8] study to reduce feature dimensionality of the vectors of the 1,000 Arabic documents. She concludes that distance-based classifier is effective and robust classifier.

Duwairi et al. [9-10] conducted those two studies in the field of Arabic text categorization. Duwairi et al. [9] studied the effect of root-based stemming and light stemming to reduce feature dimensionality of Arabic documents. They noticed that root-based stemming leads to reduce a number of semantically different Arabic words to the same three consonants (Triliteral) verbs, and leads to flaws in the process of automatic classification. Light stemming aims to

The holy Quran uses a spelling that is different from the Arabic spelling used these days in newspapers, schools, university, Arabic literature which is known as Modern Standard Arabic (MSA). Let us consider the following four Arabic words extracted from Al-Fatiha sura (Chapter 1) as shown in http://www.quranexplorer.com/quran/ and http://corpus.quran.com. The Dagger Alif or superscript Alif (دير Альф, " cavalier"
which is written as a short vertical stroke on top of an Arabic letter are used in Arabic words presented in table I. Table I exhibits clearly the difference in the spelling of holy Quran and MSA spelling for the same words.

<table>
<thead>
<tr>
<th>Sura (Chapter)</th>
<th>Ayah (Verse)</th>
<th>English meaning</th>
<th>Modern Standard Arabic (MSA)</th>
<th>Quranic Arabic</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>The Worlds</td>
<td>العالمين مالك</td>
<td>مالك العالمين</td>
</tr>
<tr>
<td>1</td>
<td>4</td>
<td>Owner</td>
<td>مالك</td>
<td>مالك</td>
</tr>
<tr>
<td>1</td>
<td>6</td>
<td>The path</td>
<td>الصراط خنجرة</td>
<td>الصراط</td>
</tr>
<tr>
<td>1</td>
<td>7</td>
<td>Path</td>
<td>الصراط خنجرة</td>
<td>الصراط</td>
</tr>
</tbody>
</table>
reduce Arabic words to their stems, and that means to remove only prefixes and suffixes from the Arabic words. Therefore in most cases light stemming would not reduce a number of semantically different Arabic words to the same stem. The K-Nearest Neighbor (K-NN) classifier was used in their study using a dataset consisting of 15,000 Arabic documents which distributed equally on three classes. Duwairi et al. [9] conclude that using light stemming yields a more accurate classification results relative to the use of Arabic root-based stemming. Duwairi et al. [10] improves their previous study [9] by adding a new feature reduction technique (word clusters) in addition to the two feature reduction techniques (root-based stemming and light stemming) already tested before. They use the same classifier (K-NN), dataset and classes used in [9]. Also they conclude that light stemming as feature reduction technique leads to a more accurate classification results relative to the other two feature reduction techniques.

Three automatic text classification algorithms (Naive Bayes (NB), K-NN, and Rocchio) were tested on a dataset consisting of 1,445 Arabic text documents by [11]. These documents of this dataset are not equally distributed on the nine categories used in their study. They conclude that NB classifier is the best, followed by K-NN and Rocchio respectively.

Associative classification (AC) is tested in the study of [7]. They test three algorithms Multi-class classification based on association rule (MCAR), NB, and SVM. A dataset consisting of 5,121 Arabic documents is partitioned equally into 7 categories. Alwedyan et al. [7] study concludes that MCAR is more accurate to automatically classify Arabic documents than the other two classifiers.

A small sample of the Arabic sayings (Hadith/Hadeeth) of Prophet Mohammed (Peace and blessings of Allah be upon him (PBUH)) is used by [13] as an Arabic dataset to find the best method that can be used to classify different Arabic sayings of the Prophet into their appropriate category. Al-Kabi et al. [13] use 80 Sayings and 12 classes. The six methods (NB, inner product, Jaccard, cosine, Dice, and Euclidean) were under test. They conclude that NB is the best method that can be used to classify different sayings (Hadith/Hadeeth) of Prophet Mohammed (PBUH) into their appropriate class (book).

Hadi et al. [6] use a dataset consisting of 2,244 Arabic documents. These documents are unequally partitioned into 5 Islamic categories. Hadi et al. [6] study aims to evaluate two automatic classification algorithms (Support Vector Machine (SVM) and (NB)), They conclude that the effectiveness of SVM is better than the effectiveness of NB.

Al-Kabi et al. [12] study is considered primitive and a limited in the field of the topical classification of the holy Quran āyāt (verses), since it does not include all the 6,236 verses of the holy Quran. It is limited to Fatihā (7 āyāt (verses)) and Yāsīn (83 āyāt (verses)). That means that their study is limited to the topical classification of only 90 āyāt (verses) out of only 6,236 āyāt (verses) consisting the whole Quran, in other words this study covers around 0.014 of the total number of the holy Quran verses. Also the number of topics used was limited to only 15.

III. METHODOLOGY

This study aims to evaluate the effectiveness of a four well-known classification algorithms (Decision Tree, K-Nearest Neighbor (K-NN), Support Vector Machine (SVM) and Naïve Bayes (NB)) to classify different Quranic āyāt (verses) according to their topics.

Our framework includes the following procedures:
1. Removal of Arabic Diacritics (‘،’،‘‘،’‘).
2. Removal of Quranic Symbols (such as: ‘،’،‘‘،’‘).
3. Using the manual human topical classification of Quranic āyāt (verses) by [14] to train and evaluate the four classifiers under consideration.

Many books classified different Quranic āyāt (verses) written by Muslim scholars. These books are different widely in their sizes. Some of these books consist of a number of volumes, and some consists of few pages. This study is based on the lexicon of Quran topics by [14]. In this lexicon the Quranic āyāt (verses) are classified into the following 14 main categories shown in table III.

<table>
<thead>
<tr>
<th>Table III: MAIN CATEGORIES OF ARABIC QURANIC āYĀT (VERSES)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Arabic Main Categories</strong></td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>3</td>
</tr>
<tr>
<td>4</td>
</tr>
<tr>
<td>5</td>
</tr>
<tr>
<td>6</td>
</tr>
<tr>
<td>7</td>
</tr>
<tr>
<td>8</td>
</tr>
<tr>
<td>9</td>
</tr>
<tr>
<td>10</td>
</tr>
<tr>
<td>11</td>
</tr>
<tr>
<td>12</td>
</tr>
<tr>
<td>13</td>
</tr>
<tr>
<td>14</td>
</tr>
</tbody>
</table>
Each of these categories is classified into subcategories. Table IV shows the seven subcategories of the first main topic (Pillars of Islam).

### IV. SUBCATEGORIES OF PILLARS OF ISLAM MAIN CATEGORY

<table>
<thead>
<tr>
<th>i</th>
<th>Arabic Subcategories of Pillars of Islam</th>
<th>English Translation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>الفصل الأول: التوحيد</td>
<td>Chapter I: Unification</td>
</tr>
<tr>
<td>2</td>
<td>الفصل الثاني: الدين</td>
<td>Chapter II: Religion</td>
</tr>
<tr>
<td>3</td>
<td>الفصل الثالث: النبوة (محمد صلى الله عليه وسلم)</td>
<td>Chapter III: The prophet Muhammad, peace be upon him</td>
</tr>
<tr>
<td>4</td>
<td>الفصل الرابع: الصلاة</td>
<td>Chapter IV: Prayer</td>
</tr>
<tr>
<td>5</td>
<td>الفصل الخامس: الصيام</td>
<td>Chapter V: Fasting</td>
</tr>
<tr>
<td>6</td>
<td>الفصل السادس: الزكاة</td>
<td>Chapter VI: Zakat and alms</td>
</tr>
<tr>
<td>7</td>
<td>الفصل السابع: الحج والعمرة</td>
<td>Chapter VII: Hajj and Umrah</td>
</tr>
</tbody>
</table>

### V. EXPERIMENTS

The experimental results for the dataset showed that the Naive Bayes (NB) yields more accurate results than the other three classifiers. In order to evaluate Arabic text Quranic āyāt (verses) classification six measures were used. The holy Quran āyāt (verses) are classified into 14 main classes as shown in Table III, and each of these is classified into a number of subclasses. Table IV presents seven subclasses of the main class (Pillars of Islam, "أركان الإسلام", "واعظة الالهة"). The three selected classes used in this study are (Ignorant of religion, "الأجاعن بالدين"), (Oneness of God, "الإسلام"), and (Penalty of Apostates, "الحجة المركون"). The class of (Oneness of God, "الإسلام") contains 20 subclasses. The subclass of (The right to God, "الحجة المركون"), contains 6 subclasses, while the subclass of (Divinity, "الإعفية"), contains 5 subclasses, where two of them; (Added Qualities, "الصفات المفردة") contain 81 subclasses, and (Individual Qualities, "الصفات المفردة") which contains 96 subclasses.

The six measurements presented in tables V, VI, VII, and VIII are True Positive (TP) rate, False Positive (FP) rate, Precision (P), Recall (R), F-measure (F1) which combines the previous two measures P and R, Receiver Operating Characteristic (ROC).

Formula 1 shows the Accuracy formula [15].

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN} \quad \text{………………(1)}$$

Formula 2 shows the Recall formula [15].

$$\text{Recall} = \frac{TP}{TP + FN} \quad \text{…………………………..(2)}$$

Formula 3 shows the Precision formula [15].

$$\text{Precision} = \frac{TP}{TP + FP} \quad \text{…………………………..(3)}$$

Where TP is the number of verses correctly classified as belonging to a class i (“true positive”). FP is the number of verses falsely classified as belonging to a class i (“false positive”) and FN is the number of verses falsely classified as not belonging to a class i (“false negative”) [15].

The three selected topical classes used in tables V, VI, VII, and VIII are:

- Class I: Includes all Quranic āyāt (verses) classified as (Ignorant of religion, "الجاعون بالدين").
- Class II: Includes all Quranic āyāt (verses) classified as (Oneness of God, "الإسلام").
- Class III: Includes all Quranic āyāt (verses) classified as (Penalty of Apostates, "الحجة المركون").

Naive Bayes (NB) classifier yields an accuracy of 99.9099%, and an error rate of 0.0901% for the dataset used in this study.

Table V shows the values of different measurements which are used to evaluate the effectiveness of Naive Bayes classifier to classify Quranic āyāt (verses) of the dataset into the three selected topical subclasses.
Table V shows that the class of (Oneness of God, "توحيد الله") gets the optimal values in the classification. While the class of (Penalty of Apostates, "جزاء المرتدين"), and the class of (Ignorant of religion, "الجاهلون بالدين"), get accepted values for the different measurements.

Table VI shows the six measurements which are used to evaluate the effectiveness of K-Nearest Neighbor (K-NN) classifier to classify Quranic āyāt (verses) of the dataset into the three selected topical classes. Using (K = 1) yields an accuracy of 99.8198%, and an error rate of 0.1802%.

Table VI shows that the effectiveness of K-Nearest Neighbor (K-NN) classifier is lower than the effectiveness of Naive Bayes. Also Table VI exhibits K-NN effectiveness is acceptable.

Table VII shows the six measurements which are used to evaluate the effectiveness of J48 Decision Tree classifier to classify Quranic āyāt (verses) of the dataset into the three selected topical classes. J48 Decision Tree classifier yields an accuracy of 99.5946%, and an error rate of 0.4054% on the dataset used in this study.

Table VIII shows that the Support Vector Machine (SVM) classifier yields lower values than the Naive Bayes, but it is more effective than K-Nearest Neighbor (K-NN) classifier.

The results of the previous tables show that the Naive Bayes classifier yields the best results to classify the Quranic āyāt (verses) belonging to the three selected classes.
VI. CONCLUSION AND FUTURE WORK

Classification of the Holy Quran is unlike typical classification. First this is since we are dealing with a sacrified script and hence classification should be context sensitive. Further, there is a need to understand existed and possible classifications for Quran verses. One of the popular Quran text classification is to classify verses into: (Meccan or Medinan). Other possible classifications that can give deeper understanding of context or semantics should be thoroughly investigated in future research related to the Holly Quran.

In this paper, we measured the effectiveness of four classification techniques ((Decision Tree, (K-NN), (SVM) and (NB)) to classify different Quranic Ayāts or verses according to their topics. We have proposed preprocessing of the classification process by performing the removal of Arabic Diacritics (‘ا, ‘ا, ‘ا, ‘ا) and removal of Quranic Symbols (such as: ١٠). We also used the manual human topical classification of Quranic Ayāts by [14] to train and evaluate the four classifiers under consideration. Three selected topics (of classes) (Ignorant of religion, "ا'), (Oneness of God, "ا''ا') and (Penalty of Apostates, "ا') according to [14] are used. This means that 1,227 Ayāts were used in this study out of total number 6236 Ayāts in the whole Quran.

Our evaluation results showed that the NB classifier scored the highest accuracy results with lowest error rate whereas the J48 Decision Tree classifier scored the lowest accuracy results with highest error rate. In the future we plan to include all topical classes and subclasses of the holy Quran. That means our future studies will include larger datasets and a larger number of classes and subclasses.

REFERENCES