The Qur'anic Classification Uses Algorithm C4.5

1st Mohamad Irfan, 2nd Wisnu Uriawan[1], 3rd Nur Lukman[1], 4th Opik Taupik Kurahman[1], 5th Wahyudin Darmalaksana[2]
{irfan.bahaf@uinsgd.ac.id[1], wisnu.u@uinsgd.ac.id[1], n.lukman@uinsgd.ac.id[1]}

UIN Sunan Gunung Djati, Department of Informatics, Bandung[1], UIN Sunan Gunung Djati, Research Center, Bandung[2]

Abstract. According to the place he descended, the surah can be divided into Makkiyah and Madaniyah groups. This division is based on the place and time that is predicted to decrease the surah or certain verses, where the surah that descended before the Prophet S.A.W migrated to Madinah classified as surah makkiyah while the surahs that descended after it belongs to Madaniyah surah. Surah that descended on Mecca in general suras with a small number of verses, contains the principles of faith and morality, the call is addressed to humans. While the surahs that descend on Medina, in general, have a large number of verses, containing the rules that govern one's relationship with God, or someone with another (shari'ah) and other discussions. The division by phase before and after the hijrah is considered more appropriate because there is a surah Madaniyah which descended on Mecca. the surah that was sent down in Mecca was 86 suras, and 28 suras were revealed in Medina. This grouping is done by determining the classification of data and using an algorithm. Algorithm C4.5 is one induction decision tree to conduct the classification process, the results of testing the accuracy of applications made using C4.5 algorithm is 95.6%. From this result, it is known that the C4.5 algorithm is quite well used in the classification process for the classification of suras in the Qur'an

Keywords: Classification; Makkiyah; Madaniyah; Algorithm C4.5

1 Introduction

The Qur'an consists of 114 surahs, each of which consists of several verses. The number of verses in the Qur'an reaches 6236 verses. The Qur'an is divided into 30 sections called juz. In studying the science of the Qur'an, must know about the classification or groupings contained in the Qur'an[1]–[3]. The importance of classification serves for further study of the clause. As we know the classification of suras in the Qur'an is divided into 2 namely surah Makkiyah and Madaniyah. the Makkiyah surah which he revealed in Mecca and the surah Madaniyah was revealed in Medina[4]–[8].

Based on the short length of the verse, the suras in the Qur'an are grouped into four groups: As Sab'uththiwaal (very long surah of 130-290 verses), Al-mi'uun (long suras of 100-129 verses), Al-matsaani (surah long enough from 51-99 verses), and Al-mufashshol (surah not long from 1-50 verses). The grouping of surah based on the short length of the verse aims to facilitate the study and study of the Qur'an[9]–[12]. Based on that, the function of Classification of surah automatically made it easier to determine the surah based on the number of verses in the Qur'an[8], [13], [14]. The determination uses Algorithm C4.5 for

ICONQUHAS 2018, October 02-04, Bandung, Indonesia

DOI 10.4108/eai.2-10-2018.2295558

grouping and forming a decision tree. The decision tree is usually expressed in tabular form with attributes and records[1], [15], [16].

This paper is organized into four sections. First, the introduction section that explains the background, problems, and objectives of the study. Second, the research methods section that describes the methods used for Algortihm C4.5. Third, finding and discussions. The last section is the conclusion.

2 Research Methods

2.1 Classification

Classification is a process for finding models or functions that explain and distinguish concepts or data classes in order to estimate the class of an object whose label is unknown. This can be said as a classification that maps an item into one of several defined classes[17], [18].

2.2 Algorithm C4.5

Algorithm C4.5 is an algorithm used to form decision tree. A decision tree is a very powerful and well known method of classification and prediction[19]–[22]. The decision tree method transforms a very large fact into a decision tree that presents the rules[1], [15], [19], [23],[16]. In general, the C4.5 algorithm for building decision trees is as follows:

1. Select the attribute as root.
2. Create a branch for each value.
3. For the case in the branch.
4. Repeat the process for each branch until all in the case have the same class.

Selecting the attribute as the root of a tree is to determine the highest gain value of the attributes. The formula for calculating the gain value of equation 1 as follows:

$$Gain\ (A) = Entropi\ (S) - \sum_{i=1}^{k} \frac{|S_i|}{|S|} \times Entropi\ (s_i)$$

Descriptions:

S : the set of cases
A : attribute
n : number of attribute partition A
|Si | : the number of cases on the i-th partition
|S| : the number of cases in S

To calculate the value of entropy can be seen in equation 2 as follows:

$$Entropi\ (S) = \sum_{j=1}^{k} -p_j * log_2\ p_j$$

Keterangan :

S : the set of cases
k : The number of S partitions
p_j : The probability of Si against S

2.3 Decision Tree

The first part of this decision tree is the root point, whereas each branch of the decision tree is a division based on test results, and the end point (leaf) is the resulting class division.

The decision tree has 3 types of nodes, namely:

1. The root node, where the dot has branches that enter and has more than one branch, sometimes has no branch at all. This node is usually an attribute that has the greatest influence on a particular class.
2. Internal node, which has only 1 incoming branch, and has more than 1 out branch.
3. A leaf node, or end node where only 1 branch entered and has no branch at all and marked that the node is a class label.

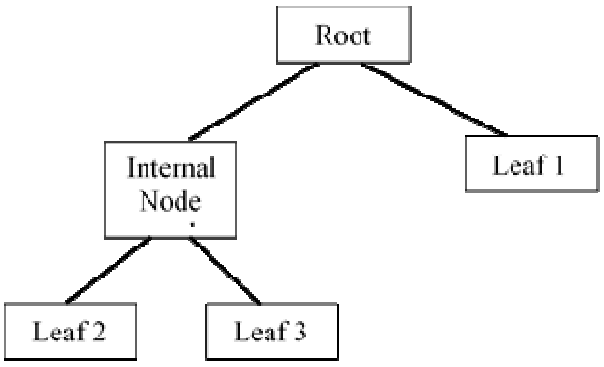

Figure 1. Tree of Decision

3 Finding and Discussions

3.1 Determining Attribute Details on Each Attribute

The following is a description of the C4.5 algorithm steps to solve the example of determining where the descendants are dropped and the number of verses based on the surah.

Table 1. Surah Sample Data Format

Attribute	Description	Range
Verses	Short	1-50 Verses
Verses	Quite Long	51-99 Verses

Verses	Long	100-129 Verses
Verses	Very long	130-290 Verses
Juz	Beginning	Juz 1-10
Juz	Midle	Juz 11-20
Juz	Last	Juz 21-30

3.2 Creating a New Case

A new case is the first step to solve a problem, by including inputs and criteria. In the final column contains an estimate of the results to be obtained later, the decision Yes (Madaniyah) and No (Makkiyah).

Table 2. New Case

No	**Verses**	**Juz**	**Discussion**	**Estimate**
1	Very long	Beginning	Law	Yes
2	Very long	Beginning	Tawheed	No
3	Long	Midle	Tawheed	No
4	Quite Long	Midle	Prophet	No
5	Quite Long	Midle	Tawheed	Yes
6	Short	Last	Prophet	Yes
7	Short	Last	Relationship	Yes
8	Very long	Beginning	Law	Yes
9	Long	Midle	Prophet	No
10	Long	Beginning	Relationship	Yes

3.3 Node Table Creation

Determine the greatest value of each class and make a root in every decision, by counting the number of cases. Then calculate and find the Entropy value of each attribute and the Gain value for each class. So that can be found the largest value and become the *root* tree.

Table 3. New Case Table

Node		Set of Cases (S)	Yes (Medina) (S1)	No (Mecca) (S2)	Entropy	Gain
1	**Total**	10	6	4	0.156	
	Verses					0.3060
	Short	2	2	0	0	
	Quite Long	2	1	1	0.150	
	Long	3	1	2	0.234	
	Very long	3	2	1	0.166	
	Juz					0.0812
	Beginning	4	3	1	0.187	
	Midle	3	0	3	0	
	Last	3	3	0	0	
	Discussion					0.1560
	Law	2	2	0	0	
	Relationship	2	2	0	0	
	Tawheed	3	1	2	0.234	
	Prophet	3	1	2	0.234	

3.4 Calculating Entropy and Gain Value

1. Calculates the value of Entropy

The following is to calculate the total entropy value

$$\textit{Entropi}\ \textbf{(Total)} = -(-6/10 \times (\log)_2\ (6/10)) + (-4/10 \times (\log)_2)$$
$$= -(-0.108) + (-0.048)$$
$$\textbf{= 0.156}$$

2. Calculate Gain value

The following calculates the Gain value of the verses:

(total,verses)

$$= 0.156 - \left(\left(\frac{2}{10} \times 0\right) + \left(\frac{2}{10} \times 0.150\right) + \left(\frac{3}{10} \times 0.234\right) + \left(\frac{3}{10} \times 0.166\right)\right) = 0.156 - 0 + 0.03$$

$$+ 0.0702 + 0.0498$$

$$= \textbf{0.3060}$$

3. Make a Decision Tree

Decision trees are created by looking at the highest Gain value results and possible decision results to support an accurate decision result.

The decision tree with the highest Gain value is a **Verses** with a total gain value of **0.3060**.

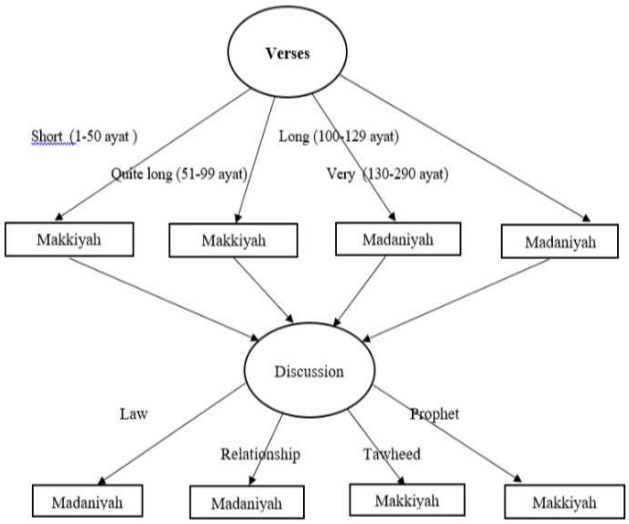

Figure 2. A decision tree with the highest Gain value

3.5 Result Classification

The following is the result of Surah Al-Quran classification using Algorithm C4.5

Table 4. New Case Table

No	Name of Surah	Number of Verses	Place	Result
	Al-Fatihah	7	Makkiyah	Al-Mufashshol
	Al-Baqarah	286	Madaniyah	As Sab'uththiwaal
	Ali Imran	200	Madaniyah	As Sab'uththiwaal
	Annisa	176	Madaniyah	As Sab'uththiwaal
	Al-Ma`Idah	120	Madaniyah	As Sab'uththiwaal
	Al-An'am	165	Makkiyah	As Sab'uththiwaal
	Al-A'raf	206	Makkiyah	As Sab'uththiwaal
	Al-Anfal	75	Madaniyah	As Sab'uththiwaal
	At-Taubah	129	Madaniyah	As Sab'uththiwaal
	Yunus	109	Makkiyah	Al-Mi'unn
	Hud	123	Makkiyah	Al-Mi'unn
	Yusuf	111	Makkiyah	Al-Mi'unn
	Arra'd	43	Makkiyah	Al-Matsaani
	Ibrahim	52	Makkiyah	Al-Matsaani
	Al-Hijr	99	Makkiyah	Al-Matsaani
	An-Nahl	128	Makkiyah	Al-Mi'unn
	Al-Isra	111	Makkiyah	Al-Mi'unn
	Al-Kahfi	110	Makkiyah	Al-Mi'unn
	Maryam	98	Makkiyah	Al-Matsaani
	Toha	135	Makkiyah	Al-Mi'uun
	Al-Anbiya	112	Makkiyah	Al-Mi'unn
	Al-Hajj	78	Madaniyah	Al-Matsaani
	Al-mu`minun	118	Makkiyah	Al-Mi'unn
	Annur	64	Madaniyah	Al-Matsaani
	Al-furqan	77	Makkiyah	Al-Matsaani
	Asy-syu'ara	227	Makkiyah	Al-Mi'unn
	An-naml	93	Makkiyah	Al-Matsaani
	Al-qashash	88	Makkiyah	Al-Matsaani
	Al-'ankabut	69	Makkiyah	Al-Matsaani
	Arrum	60	Makkiyah	Al-Matsaani
	Luqman	34	Makkiyah	Al-Mufashshol
	Assajdah	30	Makkiyah	Al-Mufashshol
	Al-Ahzab	73	Madaniyah	Al-Matsaani
	Saba`	54	Makkiyah	Al-Matsaani
	Fathir	45	Makkiyah	Al-Mufashshol
	Yasin	83	Makkiyah	Al-Matsaani
	Ash-Shaffat	182	Makkiyah	Al-Mi'unn
	Shad	88	Makkiyah	Al-Matsaani
	Az-Zumar	75	Makkiyah	Al-Matsaani
	Al-Mu`Min	85	Makkiyah	Al-Matsaani
	fushshilat	54	makkiyah	al-matsaani
	Asy-syura	53	makkiyah	al-matsaani
	Azzukhruf	89	makkiyah	al-matsaani
	Ad-dukhan	59	makkiyah	al-matsaani
	Al-jatsiyah	37	makkiyah	al-mufashshol
	al-ahqaf	35	makkiyah	al-mufashshol
	muhammad	38	madaniyah	al-mufashshol
	al-fath	29	madaniyah	al-mufashshol

	al-hujurat	18	madaniyah	al-mufashshol
	qaaf	45	makkiyah	al-mufashshol
	adz-dzariyat	6o	makkiyah	al-matsaani
	ath-thur	49	makkiyah	al-mufashshol
	Al-fil	5	makkiyah	al-mufashshol
	Al-quraisy	4	makkiyah	al-mufashshol
	Al-ma'un	7	makkiyah	al-mufashshol
	Al-kautsar	3	makkiyah	al-mufashshol
	Al-kafirun	6	makkiyah	al-mufashshol
	An-nashr	3	madaniyah	al-mufashshol
	Al-lahab	5	makkiyah	al-mufashshol
	Al-ikhlash	4	makkiyah	al-mufashshol
	Al-falaq	5	madaniyah	al-mufashshol
	An-nas	6	madaniyah	al-mufashshol

The table above is the data that can be obtained from the results of classification of the surah data according to the System. based on the number of verses, of the 114 data above shows that the results of inappropriate decisions according to the calculation of the system amounted to 5 surahs. with two comparisons of these results, we can determine how accurate this application is in determining the classification for the surah classifications, using the following formula:

Accurasy
= (mount of data-the truth from different data)/(mount of data) ×100%
= (114-5)/114 ×100%
=109/114×100 %
=95,6 %

By obtaining this number, the system of grouping surah in Al-Qur'an by using Algorithm C4.5 considered quite helpful and can facilitate determine decision by using a decision tree, so from the decision obtained the result of grouping to categorize surah in Al-Qur'an based number of verses, places of inauguration and surah.

4 Conclusion

The conclusion from the above discussion is:

1. The built system is able to categorize surah based on descended places (Makkiyah and Madaniyah), and its surah group. The system is capable of displaying the digital Qur'an and providing detailed information on each of its surahs.
2. The results of testing the accuracy of this application of 114 data surah by using Algorithm C4.5 is 95.6%, from this result that the C4.5 algorithm able to help users in grouping the suras in the Qur'an.

Suggestions related to the development of the system that has been created that is, adding new features, translating Arabic into Latin, and other content such as the suras themes and groupings of verses Muhkam and Mutasyabih in Al-Qur'an.

References

[1] D. Setiawati, I. Taufik, Jumadi, and W. Z. Budiawan, "Klasifikasi Terjemahan Ayat Al-Quran Tentang Ilmu Sains Menggunakan Algoritma Decision Tree Berbasis Mobile," *J. Online Inform.*, vol. 1, no. 1, pp. 24–27, 2016.

[2] T. A. Amal, *Rekonstruksi Sejarah Al-Quran*. Tangerang: Pustaka Alvabet, 2013.

[3] R. H. Gusmita, Y. Durachman, S. Harun, A. F. Firmansyah, H. T. Sukmana, and A. Suhaimi, "A rule-based question answering system on relevant documents of Indonesian Quran Translation," *2014 Int. Conf. Cyber IT Serv. Manag. CITSM 2014*, pp. 104–107, 2014.

[4] H. T. Sukmana, R. H. Gusminta, Y. Durachman, and A. F. Firmansyah, "Semantically annotated corpus model of Indonesian Translation of Quran: An effort in increasing question answering system performance," *Proc. 2016 4th Int. Conf. Cyber IT Serv. Manag. CITSM 2016*, 2016.

[5] E. Darwiyanto, G. A. Pratama, and S. Widowati, "Multi words quran and hadith searching based on news using TF-IDF," *2016 4th Int. Conf. Inf. Commun. Technol. ICoICT 2016*, vol. 4, no. c, 2016.

[6] D. Suryani, M. Irfan, W. Uriawan, and W. Budiawan, “Implementasi Algoritma Divide and Conquer Pada Aplikasi Belajar Ilmu Tajwid,” *J. Online Inform.*, vol. 1, no. 1, pp. 13–19, 2016.

[7] A. Zarman, M. Irfan, and W. Uriawan, “Implementasi Algoritma Ant Colony Optimization Pada Aplikasi Pencarian Lokasi Tempat Ibadah Terdekat Di Kota Bandung,” *J. Online Inform.*, vol. 1, no. 1, pp. 6–12, 2016.

[8] M. Irfan, Jumadi, W. B. Zulfikar, and Erik, “Implementation of Fuzzy C-Means algorithm and TF-IDF on English journal summary,” in *Proceedings of the 2nd International Conference on Informatics and Computing, ICIC 2017*, 2018, vol. 2018–Janua.

[9] F. Kurniawan, M. S. Khalil, M. K. Khan, and Y. M. Alginahi, “Exploiting Digital Watermarking to Preserve Integrity of the Digital Holy Quran Images,” *Proc. - 2013 Taibah Univ. Int. Conf. Adv. Inf. Technol. Holy Quran Its Sci. NOORIC 2013*, pp. 30–36, 2015.

[10] B. Hamoud and E. Atwell, “Quran question and answer corpus for data mining with WEKA,” *Proc. 2016 Conf. Basic Sci. Eng. Stud. SGCAC 2016*, pp. 211–216, 2016.

[11] E. Yosrita and A. Haris, “Identify the accuracy of the recitation of Al-Quran reading verses with the science of tajwid with Mel-Frequency Ceptral Coefficients method,” *2017 Int. Symp. Electron. Smart Devices, ISESD 2017*, vol. 2018–January, pp. 179–183, 2018.

[12] Y. M. Alginahi, O. Tayan, and M. N. Kabir, “A Zero-Watermarking Verification Approach for Quranic Verses in Online Text Documents,” *Proc. - 2013 Taibah Univ. Int. Conf. Adv. Inf. Technol. Holy Quran Its Sci. NOORIC 2013*, pp. 42–46, 2015.

[13] M. Irfan, I. Z. Mutaqin, and R. G. Utomo, “Implementation of Dynamic Time Warping algorithm on an Android based application to write and pronounce Hijaiyah letters,” in *Proceedings of 2016 4th International Conference on Cyber and IT Service Management, CITSM 2016*, 2016.

[14] M. Irfan, “Comparison between BIDE , PrefixSpan , and TRuleGrowth for Mining of Indonesian Text Comparison between BIDE , PrefixSpan , and TRuleGrowth for Mining of Indonesian Text,” *J. Phys. Conf. Ser.*, 2017.

[15] E. Elisa, “Analisa dan Penerapan Algoritma C4 . 5 Dalam Data Mining Untuk Mengidentifikasi Faktor-Faktor Penyebab Kecelakaan Kerja Kontruksi PT . Arupadhatu Adisesanti,” *JOIN (Jurnal Online Inform.*, vol. 2, no. 1, pp. 36–41, 2017.

[16] R. Revathy, “C4 . 5 Algorithm,” 2017.

[17] W. B. Zulfikar, M. Irfan, C. N. Alam, and M. Indra, “The comparation of text mining with Naive Bayes classifier, nearest neighbor, and decision tree to detect Indonesian swear words on Twitter,” in *2017 5th International Conference on Cyber and IT Service Management, CITSM 2017*, 2017.

[18] A. Wahyudin and Z. A. Hasibuan, “Research classification in strategic information system planning development: A critical review,” in *Proceedings - 2015 International Conference on Science in Information Technology: Big Data Spectrum for Future Information Economy, ICSITech 2015*, 2015, pp. 287–292.

[19] M. A. Muslim, A. Nurzahputra, and B. Prasetiyo, “Improving Accuracy of C4 . 5 Algorithm Using Split Feature Reduction Model and Bagging Ensemble for Credit Card Risk Prediction,” no. 1996, pp. 141–145, 2018.

[20] T. Sutojo, E. Mulyanto, and V. Suhartono, *Kecerdasaan Buatan*. 2011.

[21] R. S. Wahono, “A Systematic Literature Review of Software Defect Prediction : Research Trends , Datasets , Methods and Frameworks,” *J. Softw. Eng.*, vol. 1, no. 1, pp. 1–16, 2015.

[22] L. P. Irfan, Mohamad; Ayuningtias, “ANALISA PERBANDINGAN LOGIC FUZZY METODE TSUKAMOTO , SUGENO , DAN MAMDANI (STUDI KASUS : PREDIKSI JUMLAH PENDAFTAR MAHASISWA BARU FAKULTAS SAINS DAN TEKNOLOGI UNIVERSITAS ISLAM NEGERI SUNAN GUNUNG DJATI BANDUNG) Laras P : Analisa Perbandingan Logic ... L,” *J. Tek. Inform.*, vol. 10, no. 1, 2017.

[23] S. Hajrahnur, M. Nasrun, C. Setianingsih, and M. A. Murti, “Classification Of Posts Twitter Traffic Jam The City Of Jakarta Using Algorithm C4 . 5,” pp. 294–300, 2018.