

# PENERAPAN DATA MINING DENGAN ALGORITMA C4.5 UNTUK KALKULASI DATA DALAM AL-QUR'AN TERJEMAHAN

**Asmirayani**

Universitas Malikussaleh, Teknik Informatika  
Jl. Kampus Unimal BI Blang Pulo, Aceh Utara, Indonesia  
e-mail : asmirayani24@gmail.com

## **Abstract**

*Al-Qur'an was delivered by Prophet Muhammad SAW to mankind to become a guide in life in the world, which has 30 Juz, 114 suras and 6,236 verses. In the Qur'an there are many words that are repeated in a surah on a particular topic. To find out the number of repetitions of the word, an application of data calculations is built in the Al-Qur'an translation that can facilitate the calculation of words. So that users can quickly find out the number of words that repeat and how accurate the data is. This application is made with C4.5 algorithm, one of the data mining methods. Using the C4.5 algorithm is expected to be able to solve the problem of search and calculation in the translation of the Qur'an. The advantages of the C4.5 algorithm are producing decision trees that can be easily interpreted, as well as acceptable and efficient levels of accuracy. In this case the accuracy of the C4.5 algorithm reaches 80%.*

**Keywords:** al-quran, data mining, algoritma C4.5

## **Abstrak**

*Al-Qur'an disampaikan oleh Nabi Muhammad SAW kepada umat manusia untuk dijadikan pedoman dalam kehidupan di dunia, yang mempunyai 30 Juz, 114 surah dan 6.236 ayat. Dalam Al-Qur'an terdapat banyak kata yang berulang dalam suatu surah mengenai topik tertentu. Untuk mengetahui jumlah perulangan kata tersebut, maka dibangun sebuah aplikasi kalkulasi data dalam Al-Qur'an terjemahan yang dapat mempermudah pengkalkulasian kata. Sehingga pengguna bisa dengan cepat mengetahui jumlah kata yang berulang serta seberapa akuratnya data tersebut. Aplikasi ini dibuat dengan algoritma C4.5 salah satu metode data mining. Dengan menggunakan algoritma C4.5 diharapkan mampu menyelesaikan masalah pencarian dan pengkalkulasian dalam terjemahan Al-Qur'an. Kelebihan dari algoritma C4.5 yaitu menghasilkan pohon keputusan yang dapat dengan mudah diinterpretasikan, serta tingkat akurasi yang dapat diterima dan efisien. Dalam kasus ini tingkat akurasi Algoritma C4.5 mencapai 80%.*

**Kata kunci:** al-quran, data mining, algoritma C4.5

## 1. PENDAHULUAN

Sebagai kitab terakhir Al-Qur'an disampaikan oleh Nabi Muhammad SAW kepada umat manusia untuk dijadikan pedoman dalam kehidupan di dunia. Berbagai macam ilmu pengetahuan yang terdapat di dalam *Al-Quran*. Bukan hanya terpaku dalam ilmu agama, tetapi juga ilmu modern yang saat ini kita jumpai. Selain sebagai mukjizat terbesar bagi Nabi Muhammad SAW, Al-Qur'an juga membawa banyak kebaikan di dunia dan akhirat.

Al-Qur'an mempunyai 30 Juz, 114 surah dan 6.236 ayat. Dalam Al-Qur'an terdapat banyak kata yang berulang dalam suatu surah mengenai topik tertentu. Bagi sebagian orang mungkin bisa mengingat surah tersebut. Tetapi bagi orang awam, akan sulit untuk mengetahui kata tersebut terdapat di surah apa, mengingat banyaknya surah yang terdapat di dalam Al-Qur'an. Surah-surah dalam Al-Qur'an banyak yang saling berkaitan satu dengan yang lainnya. Maka diperlukannya satu kata kunci untuk mendapatkan kata yang diinginkan terletak di surah dan ayat keberapa dengan cepat, serta berapa kali kata tersebut berulang. Untuk mengetahui jumlah perulangan kata tersebut didalam terjemahan Al-Qur'an, maka dilakukan pencarian dan pengkalkulasian kata.

Ada beberapa metode yang digunakan untuk menyelesaikan masalah ini, diantaranya adalah metode algoritma C4.5. Kegunaan algoritma C4.5 adalah untuk membentuk pohon keputusan (*Decision Tree*). Adapun kelebihan dari algoritma C4.5 yaitu menghasilkan pohon keputusan yang dapat dengan mudah diinterpretasikan, serta tingkat akurasi yang dapat diterima dan efisien. Data yang digunakan oleh peneliti adalah Al-Qur'an. Untuk mempermudah proses pengkalkulasian kata dalam Al-Qur'an penulis tertarik untuk melakukan penelitian dengan judul "Penerapan *Data Mining* dengan Algoritma C4.5 untuk Kalkulasi Data Dalam Al-Qur'an Terjemahan".

## 2. METODOLOGI PENELITIAN

### 2.1 Algoritma C4.5

Algoritma C4.5 adalah algoritma yang digunakan untuk membuat pohon keputusan. Written [7] mengungkapkan bahwa Algoritma C4.5 merupakan kelompok algoritma pohon keputusan (*decision tree*). Algoritma C4.5 memiliki input berupa data *training samples* dan *samples*. *Training samples* adalah data yang digunakan untuk membangun sebuah pohon yang telah diuji kebenarannya. Sedangkan *samples* merupakan klasifikasi data.

Algoritma C4.5 dapat mengolah data numerik dan diskret. Didalam penyelesaian sebuah kasus menggunakan algoritma C4.5 ada dua elemen yang harus di pahami yaitu:

- a. *Entropy*: adalah jumlah data yang tidak relevan terhadap informasi dari kumpulan data [5].
- b. *Gain*: adalah informasi yang didapat dari perubahan entropy pada kumpulan data, baik melalui observasi atau partisipasi terhadap kumpulan data [3].

Langkah-langkah untuk membuat pohon keputusan dengan algoritma C4.5 adalah sebagai berikut [7]:

- a. Menentukan atribut untuk akar.
- b. Membuat cabang untuk masing-masing nilai.
- c. Membagi setiap kasus dalam cabang.
- d. Mengulangi proses 1 sampai 3 untuk setiap cabang, sampai semua kasus memiliki kelas.

Untuk menentukan atribut sebagai akar dilihat dari nilai *gain* tertinggi. Rumus yang digunakan untuk menghitung *gain* adalah sebagai berikut:

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy(S_i) \quad (1)$$

Keterangan:

S : Himpunan kasus

A : Atribut

n : Jumlah partisi atribut A

|S<sub>i</sub>| : Jumlah kasus pada partisi ke-i

|S| : Jumlah kasus dalam S

Sebelum menghitung nilai *gain*, terlebih dahulu menghitung nilai *entropy*. Sementara itu, perhitungan nilai *entropy* dapat dilihat pada rumus berikut:

$$Entropy(S) = \sum_{i=1}^n -p_i * \log_2 p_i \quad (2)$$

Keterangan:

S : Himpunan kasus

A : Fitur

n : Jumlah partisi S

p<sub>i</sub> : Proporsi dari S<sub>i</sub> terhadap S

### 3. HASIL DAN PERANCANGAN SISTEM

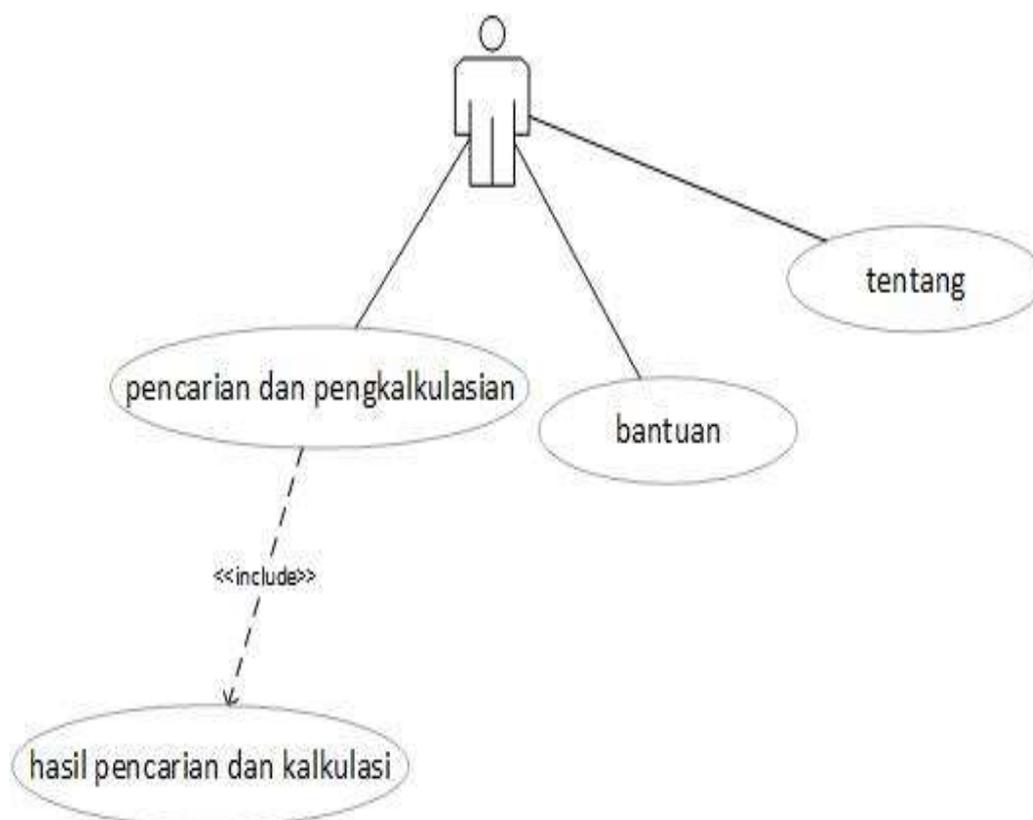
Perancangan sistem yang akan dibangun menggunakan tampilan diagram seperti *use case diagram* dan *activity diagram*. Secara garis besar, proses sistem yang akan dirancang digambarkan dengan *use case diagram* yang terdapat pada gambar berikut ini:

### 3.1. Perancangan Sistem

#### 3.1.1. Use Case Diagram

*Use case diagram* menggambarkan fungsionalitas yang diharapkan dari sebuah sistem. Yang ditekankan adalah “apa” yang diperbuat sistem, dan bukan “bagaimana”. Sebuah *use case* merepresentasikan sebuah interaksi antara *aktor* dengan sistem. *Use case* merupakan sebuah pekerjaan tertentu, misalnya login ke sistem, meng-create sebuah daftar belanja, dan sebagainya. Seorang / sebuah *aktor* adalah sebuah entitas manusia atau mesin yang berinteraksi dengan sistem untuk melakukan pekerjaan-pekerjaan tertentu[1].

Berikut rancangan *use case diagram* yang akan dibangun pada sistem dapat dilihat dari gambar:



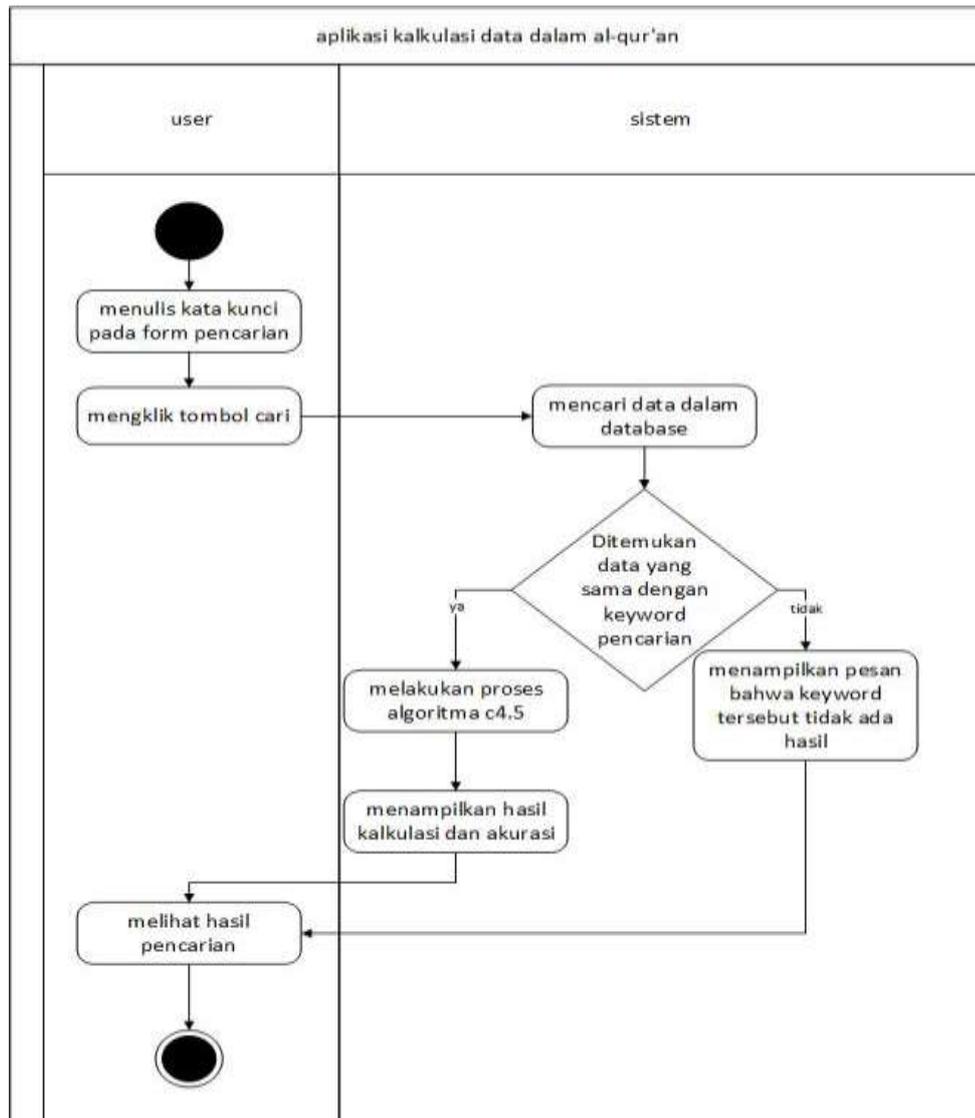
Gambar 1: *Use Case Diagram* Aplikasi Kalkulasi Data Dalam Al-Qur'an Terjemahan

#### 3.1.2 Activity Diagram

*Activity diagram* menggambarkan berbagai alir aktivitas dalam sistem yang sedang dirancang, bagaimana masing-masing alir berawal, decision yang mungkin terjadi, dan bagaimana mereka berakhir. *Activity diagram* tidak menggambarkan behaviour internal sebuah sistem (dan interaksi antar subsistem) secara eksak,

tetapi lebih menggambarkan proses-proses dan jalur-jalur aktivitas dari level atas secara umum. *Activity diagram* dapat dibagi menjadi beberapa object swimlane untuk menggambarkan objek mana yang bertanggung jawab untuk aktivitas tertentu [1].

Adapun rancangan *activity diagram* yang akan dibangun pada sistem dapat dilihat dari gambar dibawah ini:



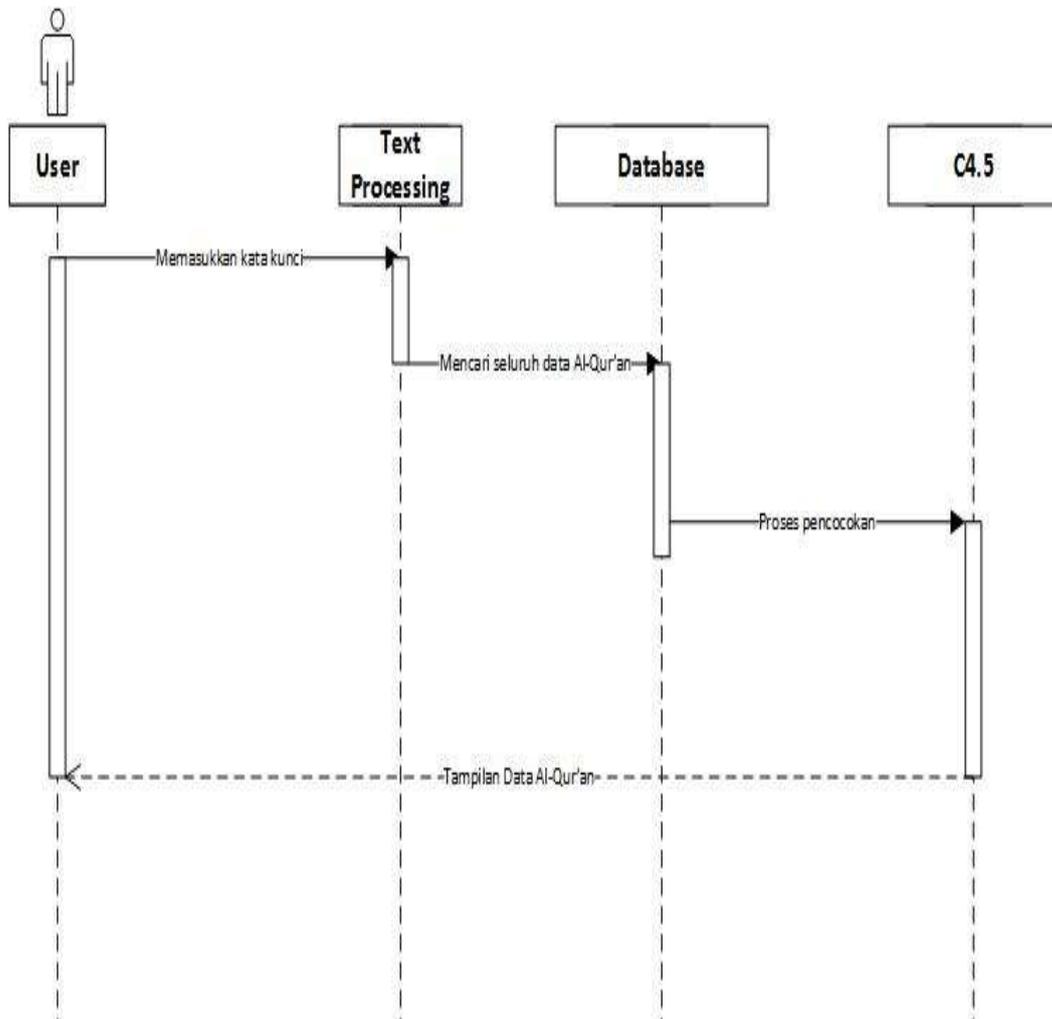
Gambar 2: Activity Diagram

### 3.1.3 Sequence Diagram

*Sequence diagram* menggambarkan interaksi antar objek di dalam dan di sekitar sistem (termasuk pengguna, display, dan sebagainya) berupa message yang digambarkan terhadap waktu. *Sequence diagram* terdiri atas dimensi vertikal (waktu) dan dimensi horizontal (objek-objek yang terkait). *Sequence diagram*

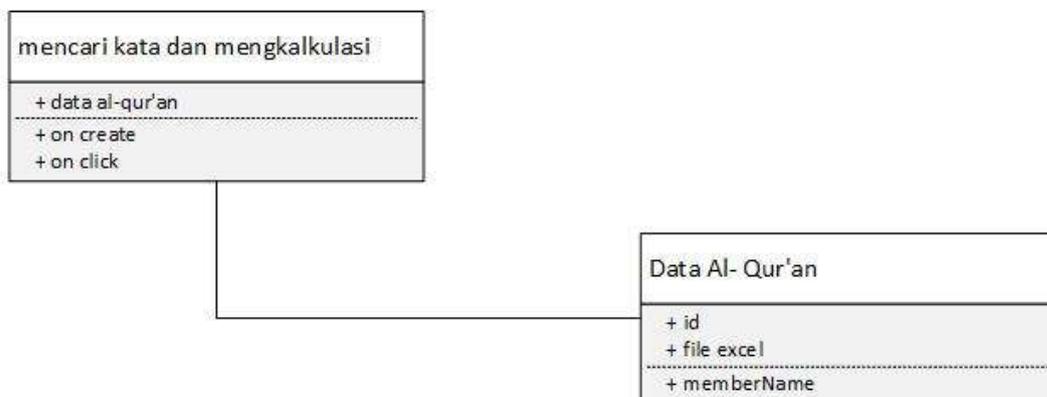
biasa digunakan untuk menggambarkan skenario atau rangkaian langkah-langkah yang dilakukan sebagai respons dari sebuah event untuk menghasilkan output tertentu. Komponen utama *sequence diagram* terdiri atas obyek yang dituliskan dengan kotak segi empat bernama *Message* diwakili oleh garis dengan tanda panah dan waktu yang ditunjukkan dengan progress vertical. Obyek atau participant diletakkan didekat bagian atas diagram dengan urutan dari kiri ke kanan. Setiap obyek terhubung dengan garis titik-titik yang disebut *lifeline*. Sepanjang *lifeline* ada kotak yang disebut *activation*. *Activation* mewakili sebuah eksekusi operasi dari obyek. Panjang kotak ini berbanding lurus dengan durasi *activation*. Sebuah *message* bergerak dari satu obyek ke obyek yang lain dan dari satu *lifeline* ke *lifeline* yang lain. Sebuah obyek dapat mengirim sebuah *message* kepada dirinya sendiri. *Time* adalah diagram yang mewakili waktu pada arah vertikal. Waktu dimulai dari atas kebawah. *Message* yang lebih dekat dari atas akan dijalankan terlebih dahulu dibanding *message* yang lebih dekat ke bawah. *Sequence diagram* menunjukkan dua dimensi, yaitu dimensi dari kiri ke kanan menunjukkan tata letak obyek dan dimensi dari atas ke bawah menunjukkan lintasan waktu [3].

Rancangan *sequence diagram* yang akan dibangun pada sistem dapat dilihat dari gambar dibawah ini:



Gambar 3: *Sequence Diagram*

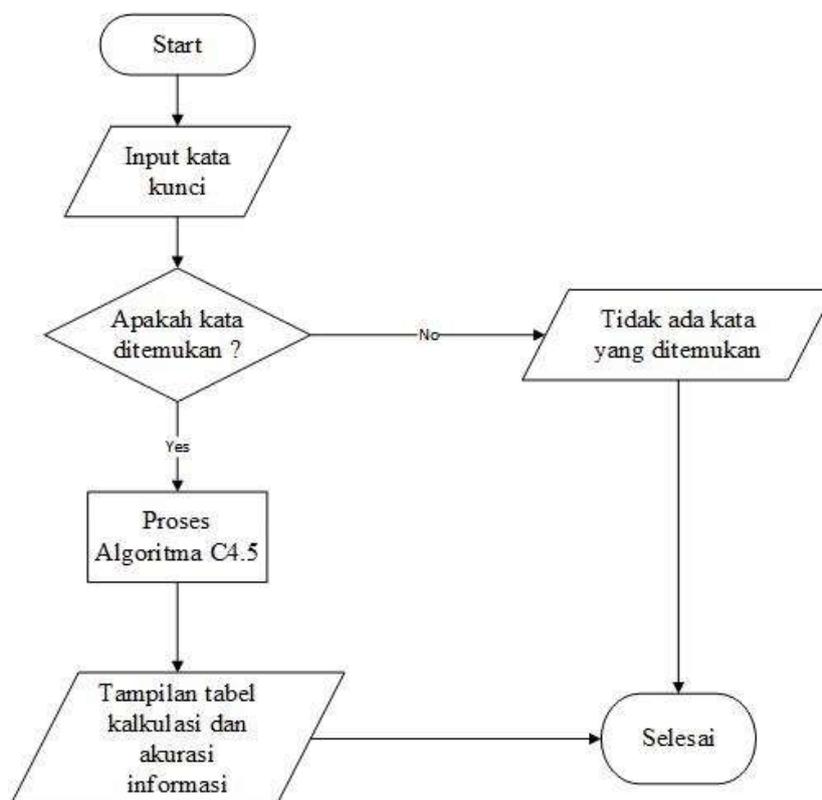
### 3.1.4 Class Diagram



Gambar 4: *Class Diagram*

### 3.1.5 Skema Sistem

Skema sistem pengkalkulasian kata dalam Al-Qur'an dengan menerapkan algoritma C4.5 dapat dilihat pada gambar berikut:



Gambar 3.2: Skema Sistem

Keterangan gambar:

- Memulai sistem
- Memasukkan kata kunci kata yang ingin dicari dan dikalkulasi
- Sistem akan memeriksa apakah kata kunci tersebut terdapat dalam data SQL.
- Jika ada maka sistem akan memproses dengan menggunakan algoritma C4.5
- Hasil proses akan ditampilkan dalam bentuk tabel kalkulasi dan akurasi informasi berdasarkan kata kunci yang dimasukkan. Sebaliknya jika kata kunci tidak ada maka konten tidak ditemukan
- Selesai

### 3.2. Hasil

Tabel 1 Data Testing

Jumlah Kata	Nama Surah	Jumlah Ayat	Persamaan Kata
63	Al-Fatihah	7	24
1.609	At-Taubah	129	60
2.466	At-Taubah	129	60
63	Al-Fatihah	7	24
1.609	Fatir	45	60
2.466	At-Taubah	129	24
1.609	Fatir	45	60
2.466	Fatir	45	24
2.466	At-Taubah	129	24
63	Al-Fatihah	7	60
1.609	Fatir	45	60
1.609	Fatir	45	24
2.466	Al-Fatihah	7	60
63	Al-Fatihah	7	24

	JML KASUS	24 (S1)	60 (S2)	ENTROPY	INFORMATION GAIN	SPLIT INFO	GAIN RATIO
TOTAL	14	7	7	1			
jumlah kata					0,163606718	1,577406283	0,10371882
	63	4	3	1	0,811278124		
	2466	5	3	2	0,970950594		
	1609	5	1	4	0,721928095		
nama surah					0,020749575	1,577406283	0,013154237
	al-fatihah	5	3	2	0,970950594		
	at-taubah	4	2	2	1		
	fatir	5	2	3	0,970950594		
jumlah ayat					0,367517645	1,046896702	0,351054354
	7	5	3	2	0,970950594		
	129	4	2	2	1		
	45	5	3	2	0,970950594		

Hasil akurasi dari nilai entropy:

Score : [100.0, 50.0, 50.0, 100.0, 100.0]

Mean Acuracy : 80.000%

#### 4. SIMPULAN

Berdasarkan hasil penelitian penulis dapat ditarik kesimpulan berupa dengan adanya aplikasi perhitungan kalkulasi data menggunakan algoritma C4.5 yang memiliki keakuratan 80% dalam pencocokan kata terjemahan Al-Qur'an, memudahkan pengguna untuk mengetahui berapa banyak pengulangan kata dalam Al-Qur'an secara cepat. Selain itu aplikasi ini dibuat secara sederhana sehingga memudahkan penggunaannya dalam menggunakan aplikasi ini.

#### DAFTAR PUSTAKA

- [1] Dharwiyanti, Sri, and Romi Satria Wahono. "**Pengantar Unified Modeling Language (UML)**." *Ilmu Komputer* (2003): 1-13.
- [2] Kusriani, Emha Taufiq Luthfi. "**Algoritma Data Mining**", Yogyakarta: Andi Offset, (2009).
- [3] Marwana. "**Algoritma C4.5 untuk Simulasi Prediksi Kemenangan Dalam Pertandingan Sepakbola, STIMED**", Nusa Palapa, Makassar, (2014).
- [4] Masduki, Yusron. "**Sejarah Turunnya Al-Qur'an Penuh Fenomenal (Muatan Nilai-Nilai Psikologi Dalam Pendidikan)**", *Medina-Te: Jurnal Studi Islam* 13.1 (2017): 39-50.
- [5] Sinaga, Jailamm Igaph, and E. Buulolo Mesran. "**Aplikasi Mobile Pencarian Kata Pada Arti Ayat Al-Qur'an Berbasis Android Menggunakan Algoritma String Matching**", *INFOTEK 2* (2016): 68-72.
- [6] Triisant. "**Pohon Keputusan dengan Algoritma C4.5**", (2015). <http://dokumen.tips/documents/algoritma-c45.html>, Diakses tgl 21 Maret 2016.
- [7] Written. Frank. "**Data Mining Practical Machine Learning tools and Techniques**", USA, (2007).