

PENDETEKSIAN MADZHAB MENGGUNAKAN TEKNIK TEXT MINING DENGAN ALGORITMA C4.5 BERBASIS WEB

Muhammad Fikry¹, Defry Hamdhana², Sri Rahayu³
Universitas Malikussaleh , Teknik Informatika
Jl. Kampus Unimal BI Blang Pulo, Aceh Utara, Indonesia
e-mail : muh.fikry@unimal.ac.id, defryhamdhana@unimal.ac.id
sr.ayu97@gmail.com*

Abstract

Madzhab is an opinion about a law from the Mujtahid Imams which we make as a guide in worshipping Allah SWT. In his opinion, the Mujtahid Imams have different ideas from each other, such as Imam Shafi'i, Hambali Imam, Hanafi Imam and Imam Maliki. To find out the differences of opinion of the Imams, an application was built that was able to display these differences from various files with word detection. Word detection is done by the text mining method, namely the C4.5 algorithm. The value obtained from the C4.5 algorithm in detecting madhab is 60%.

Keywords: madzhab, text mining, C4.5 algorithm

Abstrak

Madzhab merupakan pendapat tentang suatu hukum dari para Imam Mujtahid yang kita jadikan sebagai pedoman dalam beribadah kepada Allah Swt. Dalam berpendapat, para Imam Mujtahid memiliki pemikiran yang berbeda satu sama lain, seperti halnya Imam Syafi'i, Imam Hambali, Imam Hanafi dan Imam Maliki. Untuk mengetahui perbedaan pendapat dari para Imam tersebut, dibangunnya sebuah aplikasi yang mampu menampilkan perbedaan-perbedaan tersebut dari berbagai file dengan pendeteksian kata. Pendeteksian kata tersebut dilakukan dengan metode text mining yaitu algoritma C4.5. Nilai yang didapatkan dari algoritma C4.5 dalam pendeteksian madzhab adalah 60%.

Kata kunci: madzhab, text mining, algoritma C4.5

1. PENDAHULUAN

Perkembangan hukum Islam setelah wafatnya Rasulullah SAW berkembang sangat pesat. Hukum Islam sendiri bermula dari Al-Qur'an dan Hadist. Tetapi dengan keterbatasan ilmu serta pengetahuan kita dalam memahami maksud dari Al-Qur'an dan Hadist, kita memerlukan pendapat dari para Imam Mujtahid. Para Imam Mujtahid pun memiliki pendapat yang berbeda-beda dalam menetapkan suatu hukum. Dari perbedaan pendapat inilah, terbentuknya kelompok-kelompok

fiqh yang biasa disebut Madzhab. Namun dari banyaknya madzhab yang ada, maka hanya ada beberapa madzhab saja yang mampu bertahan hingga sekarang. Pada penelitian ini, penulis hanya membahas madzhab yang tetap eksis sampai kini, terutama madzhab-madzhab yang berkembang di Indonesia seperti: Madzhab Hanafi, Madzhab Hambali, Madzhab Syafii dan Madzhab Maliki.

Dalam madzhab tersebut akan dijelaskan tentang ilmu fiqh. Fiqh itu sendiri merupakan ilmu-ilmu yang menjelaskan tentang wajib dan fardhu atas tiap-tiap muslim seperti tentang bersuci, syarat-syarat sahnya shalat, halal haram, dan lain-lain. Seperti halnya shalat, terdapat perbedaan di masing-masing Madzhab yang empat. Dalam mazhab Hanafi ada 18 rukun dalam shalat, Madzhab Maliki dan Hambali ada 14 rukun dalam shalat dan dalam Madzhab Syafi'i menyebutkan bahwa ada 13 rukun dalam shalat. Apabila kita mencari perbedaan tersebut di Internet, akan memakan waktu yang cukup lama. Maka dibutuhkannya sistem yang mampu menampilkan perbedaan-perbedaan tersebut dari berbagai *file* dengan pendeteksian kata. Pendeteksian kata yang tepat dan sesuai kebutuhan menjadi sangat penting karena akan memudahkan dalam pencarian *file* yang dibutuhkan. Dalam hal ini penulis menggunakan Algoritma *C4.5*.

Algoritma *C4.5* adalah algoritma yang digunakan untuk membentuk suatu pohon keputusan (*Decision Tree*). Pohon keputusan ini nantinya berguna untuk *mengeksplorasi* data, serta menemukan adanya hubungan tersembunyi diantara sejumlah calon variabel *input* dengan suatu variabel target. Kelebihan dari algoritma *C4.5* ini sendiri yaitu dapat menghasilkan pohon keputusan yang mudah diinterpretasikan dan memiliki tingkat keakurasian yang dapat diterima serta efisien. Algoritma *C4.5* memiliki tingkat akurasi sebesar 7,81% dibandingkan algoritma Naïve Bayes yang dilakukan pada penelitian *Classification of Hadith Levels Using Data and Text Mining Techniques* [1]. Data yang nantinya akan digunakan oleh peneliti adalah *file pdf*. Untuk mempermudah proses pendeteksian madzhab dalam *file pdf* penulis tertarik untuk melakukan penelitian dengan judul "Pendeteksian Madzhab menggunakan Teknik *Text Mining* dengan Algoritma *C4.5* Berbasis Web".

2. METODOLOGI PENELITIAN

2.1. Algoritma *C4.5*

Algoritma *C4.5* adalah algoritma yang digunakan dalam melakukan proses klasifikasi dari suatu data dengan menggunakan teknik pohon keputusan. Algoritma *C4.5* ini merupakan lanjutan dari algoritma ID3 yang menggunakan prinsip decision tree yang mirip. Algoritma ini sangat terkenal serta banyak disukai karena memiliki cukup banyak kelebihan. Kelebihannya yaitu dapat

mengolah suatu data numerik dan diskret dan dapat menangani nilai suatu atribut yang hilang, serta mampu menghasilkan aturan yang mudah diinterpretasikan dan performa dari algoritma ini merupakan salah satu yang tercepat dibandingkan dengan algoritma-algoritma lain [6].

Dasar dari algoritma ini adalah ide pembuatan pohon keputusan yang didasarkan dari pemilihan suatu atribut yang memiliki prioritas tertinggi atau dapat juga disebut mempunyai nilai gain tertinggi yang didapat dari nilai entropy suatu atribut tersebut sebagai patokan atribut klasifikasi [3]. Kemudian secara rekursif cabang-cabang pohon diperluas sehingga seluruh pohon terbentuk. Berdasarkan kamus IGI Global (*International Publisher of Progressive Academic*), entropy adalah jumlah dari suatu data yang tidak relevan terhadap informasi dari suatu kumpulan data[7].

Gain adalah informasi yang didapat dari nilai entropy pada suatu kumpulan data, bisa melalui observasi atau bisa juga disimpulkan dengan melakukan partisipasi terhadap suatu kumpulan data [3]. Berdasarkan kutipan dari Jefri [5], ada empat langkah untuk melakukan proses pembuatan suatu pohon keputusan pada algoritma *C4.5*, yaitu:

- a. Menentukan suatu atribut yang akan dijadikan sebagai akar.
- b. Membuat cabang untuk tiap-tiap nilai.
- c. Membagi setiap kasus yang ada kedalam cabang.
- d. Mengulangi proses untuk setiap cabang sampai semua kasus yang ada didalam cabang memiliki kelas yang sama.

Menurut Jiandi [3] data yang akan diteliti harus dijadikan sebuah tabel yang didasarkan dari kasus dan jumlah dari responden sebelum dilakukannya perhitungan untuk mencari nilai dari *entropy* dan *gain*.

$$Entropy(S) = \sum_{i=0}^n - p_i * \log_2 p_i \quad (1)$$

Rumus (1) merupakan rumus yang digunakan dalam menghitung nilai *entropy* yang akan digunakan untuk menentukan seberapa informatifkah atribut tersebut. Berikut keterangannya :

- S* : Himpunan dari suatu kasus
n : Jumlah dari partisi
p_i : Jumlah kasus yang ada pada partisi ke-i

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy(S_i) \quad (2)$$

Rumus (2) merupakan rumus yang digunakan dalam menghitung nilai *gain* setelah dilakukannya perhitungan nilai *entropy*. Berikut keterangannya :

- S : Himpunan dari suatu kasus
- n : Jumlah partisi dari atribut A
- $|S_i|$: Jumlah kasus yang ada pada partisi ke- i
- $|S|$: Jumlah kasus yang ada dalam S

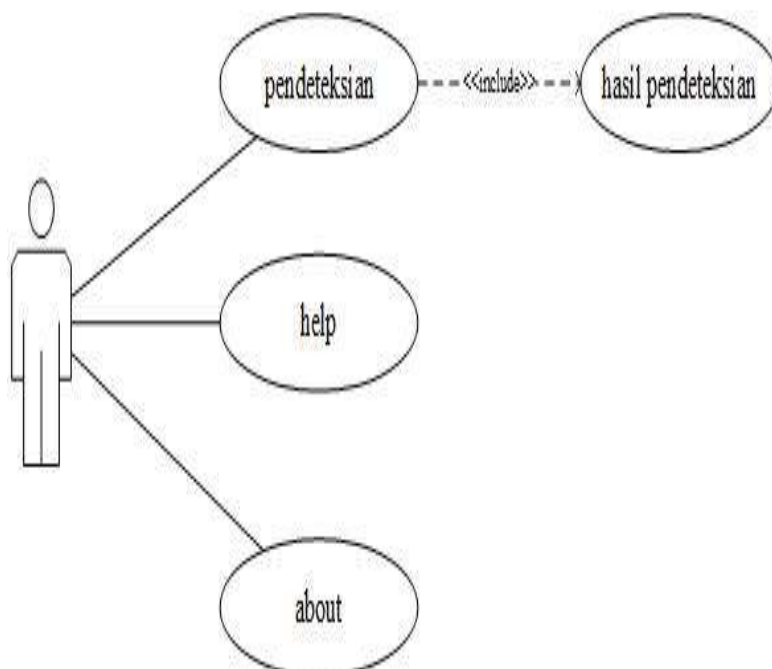
Dengan mengetahui rumus-rumus diatas, data yang telah diperoleh dapat dimasukkan dan diproses dengan algoritma *C4.5* untuk proses pembuatan decision tree.

3. HASIL DAN PEMBAHASAN

3.1 Pembahasan

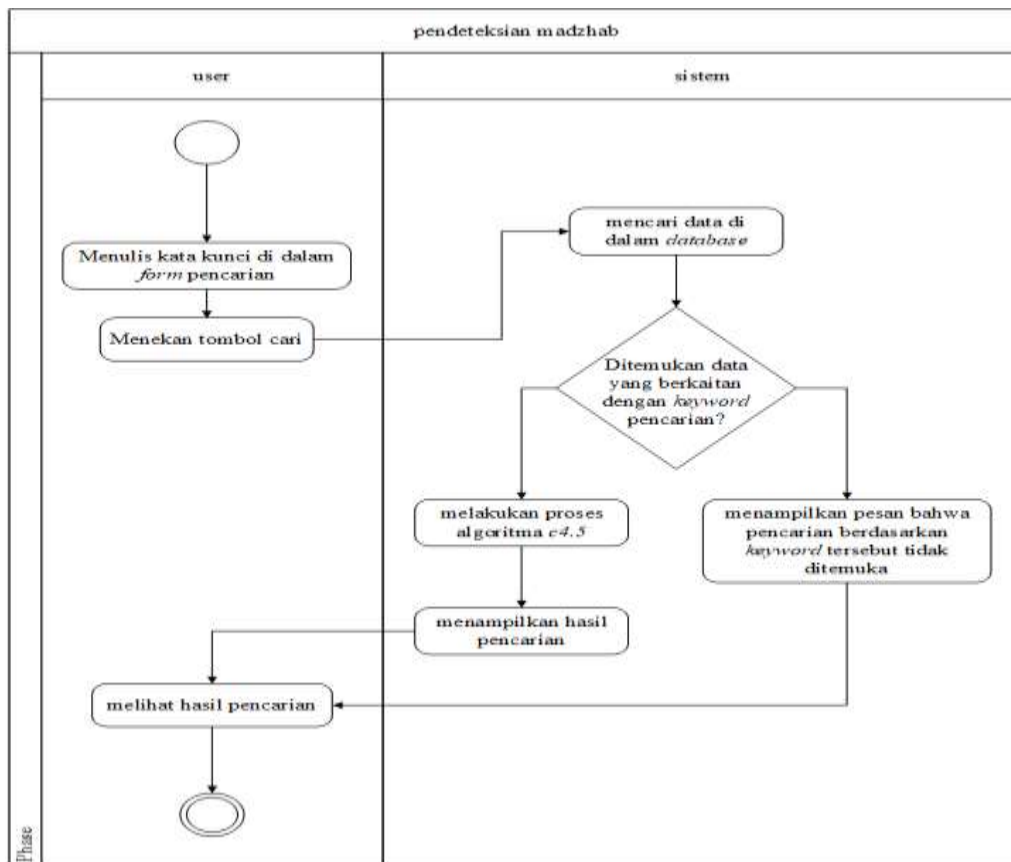
Adapun perancangan sistem yang akan dibangun yaitu analisa sistem dengan menggunakan tampilan diagram seperti *use case diagram* dan *activity diagram*.

3.1.1. Use Case Diagram



Gambar *Use Case Diagram* Aplikasi Pendeteksian Madzhab

3.1.2. Activity Diagram

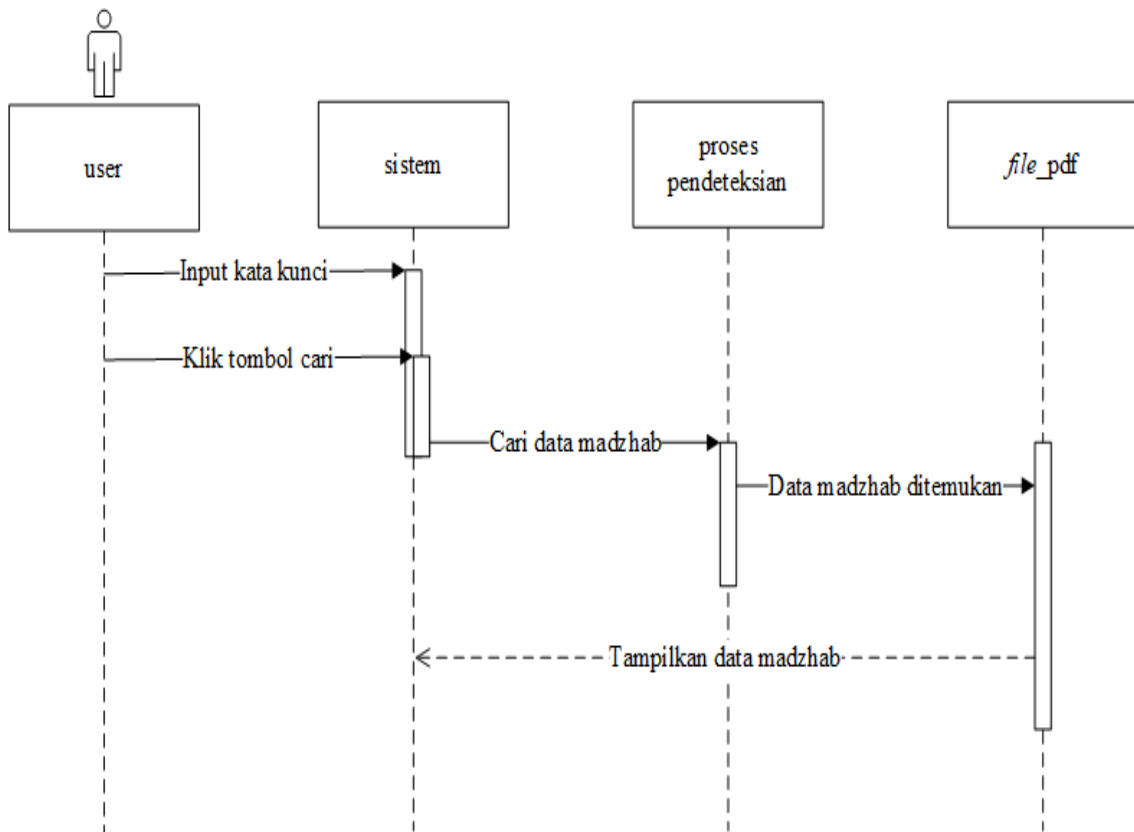


Gambar Activity Diagram Aplikasi Pendeteksiian Madzhab

Saat memulai sistem, *user* menginputkan kata kunci dan menekan *buoton* cari yang kemudian dibaca oleh sistem. Kemudian sistem akan melakukan pencarian data pada *database* sesuai dengan kata kunci yang diinputkan *user*. Jika data yang dicari ada pada *database* maka akan dilakukannya klasifikasi dengan algoritma C4.5, setelah hasil didapatkan maka sistem akan menampilkan datanya kepada *user*. Jika data yang dicari sesuai dengan kata kunci tidak ditemukan, maka sistem akan menampilkan kepada *user* bahwa data tidak ada/ tidak ditemukan.

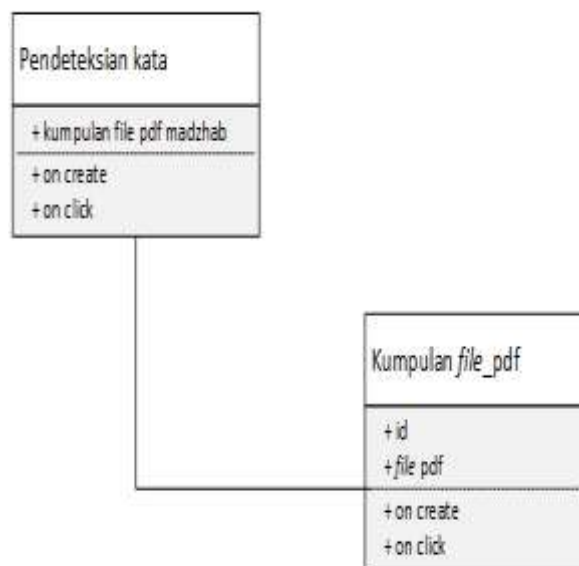
3.1.3 Sequence Diagram

Pada *Sequence Diagram* *user* akan menginputkan kata kunci dan mengklik *button* cari, kemudian sistem akan menerima perintah tersebut lalu melakukan proses algoritma C4.5 pada *file-file* pdf yang ada pada *database*. Setelah proses selesai maka sistem akan menampilkan hasil dari pencarian tersebut.



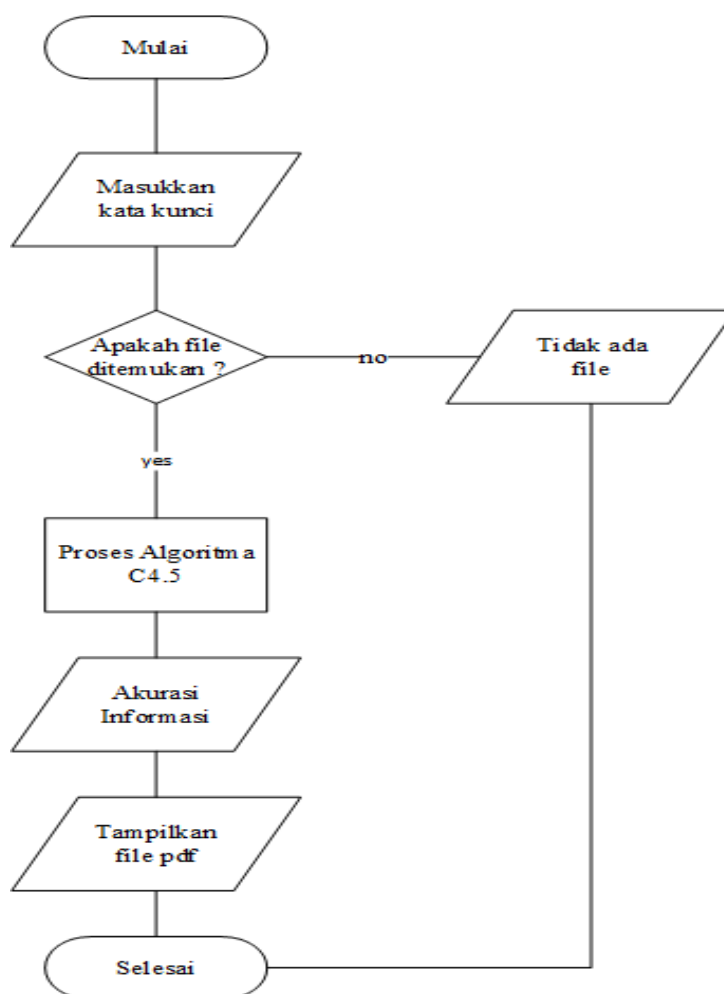
Gambar *Sequence Diagram* Aplikasi Pendeteksian Madzhab

3.1.4 Class Diagram



Gambar *Class Diagram* Aplikasi Pendeteksian Madzhab

3.1.5 Skema Sistem



Gambar Skema Sistem Aplikasi Pendeteksian Madzhab

Keterangan dari gambar skema sistem:

- Memulai sistem.
- Memasukkan kata kunci yang diinginkan. Sistem akan memeriksa apakah *file* yang sesuai dengan kata kunci tersedia atau tidak.
- Jika kata kunci yang dimasukkan tidak sesuai maka konten tidak ditemukan.
- Sebaliknya jika kata kunci sesuai maka sistem akan memproses *file* menggunakan algoritma *C4.5*.
- Hasil proses akan ditampilkan berupa akurasi informasi akan yang sesuai kata kunci dan menampilkan *file pdf* nya
- Selesai

3.2 Hasil

3.2.1. Data Testing

File Pdf	Syafi'i	Hambali	Hanafi	Maliki	Kesamaan Kata
Shalat	Ada	Ada	Tidak Ada	Ada	50
Nikah	Tidak Ada	Ada	Tidak Ada	Ada	80
Shalat	Tidak Ada	Tidak Ada	Ada	Ada	80
Haji	Ada	Tidak Ada	Ada	Ada	50
Puasa	Ada	Tidak Ada	Tidak Ada	Ada	80
Puasa	Ada	Ada	Tidak Ada	Ada	50
Warisan	Tidak Ada	Ada	Tidak Ada	Ada	50
Nikah	Tidak Ada	Ada	Tidak Ada	Tidak Ada	80
Talak	Ada	Ada	Tidak Ada	Ada	50
Tayamum	Tidak Ada	Ada	Tidak Ada	Ada	50
Tayamum	Ada	Tidak Ada	Ada	Tidak Ada	50
Warisan	Tidak Ada	Tidak Ada	Ada	Ada	80
Warisan	Tidak Ada	Ada	Tidak Ada	Ada	50
Shalat	Tidak Ada	Tidak Ada	Tidak Ada	Ada	50

Gambar Data Testing untuk Pendeteksian Madzhab

		JML KASUS	50 (S1)	80 (S2)	ENTROPY	INFORMATION GAIN
TOTAL		14	9	5	0,940285959	
FILE PDF						0,600651137
	SHALAT	3	2	1	0,918295834	
	NIKAH	2	0	2	0	
	PUASA	2	1	1	1	
	HAJI	1	1	0	0	
	WARISAN	3	2	1	0,918295834	
	TALAK	1	1	0	0	
	TAYAMUM	2	2	0	0	
SYAFI'I						-0,303279008
	ADA	6	5	1	0,650022422	
	TIDAK ADA	8	4	4	1	
HAMBALI						0,04812703
	ADA	8	6	2	0,811278124	
	TIDAK ADA	6	3	3	1	
HANAFI						0,025078174
	ADA	4	2	2	1	
	TIDAK ADA	10	7	3	0,881290899	
MALIKI						-6,509775004
	ADA	12	8	4	0,918295834	
	TIDAK ADA	2	1	1	1	

Gambar Hasil Pencarian Nilai Entropy dan Information Gain

Pencarian tersebut dilakukan pada setiap atribut yang ada. Setelah hasil entorpy didapatkan maka selanjutnya kita dapat mencari nilai gain. Untuk tingkat keakurasiannya sendiri, didapatkan dari hasil entropy pada setiap atribut yang ada.

Untuk mendapatkan tingkat akurasi dari nilai entropy dilakukan dengan menggunakan bahasa python, sehingga diperoleh sebesar 60%.

Hasil Akurasi dari nilai Entropy

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

Mean Acuracy : 60.000%

4. SIMPULAN

Walaupun sama-sama menentukan hukum berdasarkan Al-Qur'an dan Hadist, tidak menutup kemungkinan adanya perbedaan dari masing-masing madzhab. Dengan perbedaan itulah, kita mempunyai pembendaharaan pengetahuan ilmu yang sangat banyak tentang hukum Islam dan menentukan Madzhab mana yang akan kita ikuti. Dengan adanya Pendeteksian Madzhab menggunakan algoritma c4.5, dapat mempermudah kita untuk mengetahui

perbedaan madzhab tersebut serta mempercepat dalam pencarian perbedaan hukum-hukum dari masing-masing madzhab. Keakuratan algoritma C4.5 dalam kasus ini memiliki tingkat akurasi yang tinggi dengan output 60%.

DAFTAR PUSTAKA

- [1] Bustami, & Fikry, M. (2018). **Classification of hadith levels using data and text mining techniques**. Eurasian Journal of Analytical Chemistry, 13(6), 186-191.
- [2] Han, Jiawei and M. Kamber. 2001. **"Data Mining: Concepts and Techniques. USA: Academic Press."**
- [3] Marwana, 2014, **"Algoritma C4.5 untuk Simulasi Prediksi Kemenangan Dalam Pertandingan Sepakbola"**, STIMED, Nusa Palapa, Makassar.
- [4] Munawar. (2005). **"Pemodelan Visual dengan UML. Graha Ilmu"**, Yogyakarta, 17-100.
- [5] Slamet, A.,2007, **"Manajemen Sumber Daya Manusia"**, Universitas Negeri Semarang, Semarang.
- [6] Tjahyono, A. dan Anggara, A. M., 2010, **"Sistem Pendukung Keputusan Penerimaan Pegawai Baru pada PT. Kanasritex Semarang"**, Techno.com, Vol. 9 No.3.
- [7] Triisant, 2015, **"Pohon Keputusan dengan Algoritma C4.5"**, <http://dokumen.tips/documents/algoritma-c45.html>, Diakses tgl 21 Maret 2016.
- [8] Yanggo, Huzaemah Tahido, 1997, **"Pengantar Perbandingan Mazhab"**, Jakarta: Logos, Cet., I.