

ALGORITMA C4.5 DENGAN *PARTICLE SWARM OPTIMIZATION* UNTUK KLASIFIKASI LAMA MENGHAFAI AL-QURAN PADA SANTRI MAHADUL QURAN

Firman Santoso¹, Abdul Syukur², A. Zainul Fanani³

^{1,2,3}Fakultas Ilmu Komputer, Universitas Dian Nuswantoro

Email: ¹pienkman.22@gmail.com, ²abah.syukur01@gmail.com, ³a.zainul.fanani@dsn.dinus.ac.id

ABSTRAK

Pondok pesantren merupakan salah satu pendidikan yang fokus terhadap bidang keagamaan. Namun, seiring berjalannya waktu pesantren di Indonesia terus berkembang sangat pesat, sudah banyak pesantren yang didalamnya sudah dilengkapi dengan ilmu umum tidak hanya agama saja. Penelitian ini bersumber dari Pondok Pesantren Salafiyah Syafi'iyah di Asrama Ma'hadul Qur'an. Dalam penelitian ini bertujuan untuk mengklasifikasi lama menghafal alqur'an dengan menggunakan algoritma C4.5 dengan Particle Swarm Optimization (PSO). Dengan menggunakan dataset lulusan santri Ma'hadul Qur'an. Dari hasil eksperimen yang dilakukan menghasilkan Decision Tree C4.5 dengan akurasi 80 %. Setelah dilakukan dengan menggunakan C4.5 dan Particle Swarm Optimization (PSO) akurasi meningkat menjadi 87 %.

Kata Kunci : C4.5, Particle Swarm Optimization (PSO), Menghafal Al-qur'an, Salafiyah Syafi'iyah

1. PENDAHULUAN

1.1. Latar Belakang

Pondok Pesantren yang berdiri di Sukorejo, masyarakat mengenalnya dengan Pesantren Salafiyah Syafi'iyah sebelumnya adalah hutan lebat yang di huni banyak binatang buas dan makhluk halus. Setelah mendapatkan saran dari Habib Musawa dan Kiai Asadullah dari Semarang [1]. Raden Ibrahim yang kemudian dikenal dengan KHR. Syamsul Arifn, segera membat hutannya belantara tersebut pada tahun 1908 dengan dibantu oleh saudaranya, KH. Abd. Latif [2]. Dipilihnya hutan yang terletak di sebelah timur kabupaten Panarukan (sekarang menjadi Kabupaten Situbondo), berdasarkan hasil istikharah. Dan pada tahun 1914-1923, KHR. Syamsul Arifin berhasil mendirikan Pondok Pesantren kecil dan menjadi pendiri dan pengasuh pertama.

Di pesantren tersebut seluruh santri menempuh pendidikan baik formal ataupun nonformal. Beberapa pendidikan formal yang ada diantaranya Madrasah Ma'hadul Qur'an, Madrasah Ibtidaiyah sampai Madrasah Aliyah, dengan mata pelajaran murni tentang keagamaan. Dan jenjang pendidikan tingkat SD sampai dengan Perguruan Tinggi, dengan waktu pelaksanaan setelah sekolah Madrasah. Pendidikan Nonformal diantaranya, lembaga pengembangan bahasa asing, lembaga qiro'tuna amsilatuna dan lain-lain.

Al qur'an dan *kitab kuning*, merupakan pelajaran pokok yang harus dikuasi setiap santri. Sehingga, dua bidang tersebut menjadi salah satu persyaratan lulus disetiap jenjang pendidikan. Al qur'an merupakan salah satu dari lima wasiat pendiri dan pengasuh kedua, KHR. As'ad Syamsul Arifin. Dan setiap santri baru, diwajibkan mengikuti kegiatan ekstrakurikuler untuk pendalaman al qur'an tingkat dasar, pada lembaga qiro'atuna.

Bagi santri yang berkonsentrasi untuk menghafal al qur'an, berada di lembaga *Ma'hadul Qur'an* (MQ). Selama 24 jam, santri yang mengenyam pendidikan di MQ selalu mendapatkan pengawasan ketat dalam hal menghafal Al Qur'an. Sehingga, waktunya pun diatur dari kegiatan menghafal hingga proses menyelesaikan hafalan 30 juz.

Namun tidak semua santri MQ berhasil menghafal Al qur'an dengan cepat. Sebab, dalam memperoleh tingkatan hafalan dengan baik dan benar tidak hanya cukup dengan menghafal sekali saja. Metode mengulang-ulang (*Takrar*) merupakan salah satu pendukung untuk bisa menghafal dan mengingat-ingat kembali ayat-ayat yang dihafal [3]. Minimnya kesibukan dan tingkatan usia juga salah satu penunjang kemudahan al-qur'an, siswa MQ tidak semua hanya terfokus pada satu pendidikan, ada yang merangkap pendidikan sore yakni tingkat SD sampai tingkat Perguruan Tinggi. Ada pula yang hanya fokus pada pendidikan MQ saja. Namun selain pendidikan, usia pun juga menjadi penunjang untuk menghafal.

Dari berbagai sumber yang di dicari, sejauh ini belum ditemukan penelitian tentang klasifikasi lama menghafal al-qur'an. Namun, penelitian yang mengangkat topik klasifikasi data mining dengan metode C4.5 sudah banyak yang menggunakannya. Diantaranya yang dilakukan oleh Sukardi, Abd. Syukur, dan Catur Supriyanto tentang Klasifikasi Spam email menggunakan algoritma C4.5 dengan Seleksi Fitur, Eksperimennya menghasilkan klasifikasi yang masuk kategori sangat baik. Sebab algoritma C4.5 dengan menggunakan tiga model Gain Ratio, Information gain, dan gini index. Menghasilkan nilai akurasi yang tinggi pada model kriteria gini index yakni 92,18%, dari hasil tersebut kemudian dilakukan seleksi fitur menggunakan chi-square, information gain, information gain ratio. Dan ditemukan hasil akurasi yang tinggi pada information gain ratio dengan nilai $p=0,6$ dengan hasil akurasi menjadi 92,46% dan memiliki nilai AUC rata-rata antara 0,9 – 1,0 [4].

Pemanfaatan algoritma C4.5 pernah dilakukan oleh Liliana Swastina dalam meneliti penentuan jurusan mahasiswa, Tujuannya menggunakan algoritma C4.5 yang dibandingkan dengan Naive Bayes diharapkan mampu menjadi alat pendukung keputusan dalam proses penentuan jurusan mahasiswa oleh pihak perguruan tinggi. Dari hasil uji yang dilakukan pada penelitiannya, algoritma Decision Tree C4.5 lebih akurat diterapkan untuk penentuan kesesuaian jurusan mahasiswa dengan hasil tingkat keakuratan 93,31% dan menghasilkan akurasi rekomendasi jurusan sebesar 82,64% [5].

Dalam perumusan semacam ini, Zhiwu Liu dkk [6] telah memanfaatkan data mining untuk melakukannya. Mereka memanfaatkan pola prediksi yang dihasilkan pohon keputusan. Algoritma data mining yang dapat digunakan untuk pembentukan pohon keputusan sangat banyak, diantaranya: Algoritma ID3, CART, dan C4.5. Algoritma ID3 dikembangkan menjadi Algoritma C4.5 [7].

Klasifikasi merupakan salah satu teknik pada data mining [8]. Dan dalam klasifikasi harus ada variabel kategori target [7]. Ada banyak algoritma pelatihan yang bisa digunakan pada klasifikasi, diantaranya Decision Tree, K-Nearest Neighbor, Artificial Neural Network, Support Vector Machine, dan sebagainya [9]. Mengacu pada pertimbangan sebelumnya, pendekatan data mining dengan menerapkan Algoritma *Decision Tree* C4.5 akan dilakukan untuk mengklasifikasi lamanya menghafal al-qur'an pada santri MQ. Sebab, metode klasifikasi dengan menggunakan *Decision Tree* C4.5 memiliki teknik pohon keputusan yang populer, kefleksibelan membuat metode ini interaktif, khususnya karena memberikan keuntungan berupa visualisasi saran (dalam bentuk decision tree) yang membuat prosedur prediksinya dapat diamati [9].

Keunggulan dari algoritma C4.5 adalah memiliki kemampuan dalam mengolah data set seperti kecepatan klasifikasi, setiap atribut bersifat diskrit, binari dan kontinue, serta transparansi pengetahuan atau klasifikasi [10]. Seperti halnya algoritma yang lain, Algoritma C4.5 juga memiliki kelemahan yang terletak pada pembacaan data ketika berjumlah besar. Sehingga untuk meningkatkan kinerja klasifikasi Algoritma C4.5 perlu ditambahkan algoritma yang lain. Dalam kasus ini, Algoritma *Particle Swarm Optimization* (PSO) sebagai fitur seleksi lebih efektif karena hanya membutuhkan parameter sedikit waktu komputasi. Sehingga, dalam penelitian Meng-Chang Tsai Dkk. Menjelaskan, bahwa PSO digunakan untuk meningkatkan kinerja klasifikasi [11].

1.2. Rumusan Masalah

- a. Adanya kesulitan dalam menentukan strategi metode hafalan al-qur'an pada santri Ma'hadul Qur'an.
- b. Belum ditemukan peneliti yang mengangkat topik tentang klasifikasi lama menghafal al-qur'an.

1.3. Tujuan Penelitian

Dengan mengacu pada perumusan masalah, maka tujuan yang ingin dicapai pada penelitian ini adalah sebagai berikut :

- a. Memudahkan dalam menentukan strategi metode hafalan al-qur'an pada santri Ma'hadul Qur'an.
- b. Membuat suatu model klasifikasi lama menghafal al-qur'an pada santri MQ menggunakan algoritma *decision tree C4.5* dengan *Particle Swarm Optimization (PSO)* dengan nilai akurasi yang baik.

1.4. Manfaat Penelitian

Manfaat dari penelitian ini adalah :

- a. Bagi iptek
Sebagai kontribusi dalam bidang ilmu komputer serta diketahuinya akurasi *Algoritma C4.5* dengan *Particle Swarm Optimization (PSO)* yang memiliki kinerja paling baik.
- b. Bagi Lembaga terkait
Bisa memberikan informasi terbaik dari hasil klasifikasi lamanya menghafal al-qur'an serta sebagai kontribusi dalam menyelesaikan masalah yang ada di lembaga dari hasil penelitian yang dilakukan mampu memberikan hasil yang mendekati kenyataan.

2. TINJAUAN PUSTAKA

2.1. Penelitian Terkait

Penelitian terdahulu yang berkaitan tentang klasifikasi dengan menggunakan algoritma C4.5 dan Particle Swarm Optimization pernah dilakukan oleh Meng-Chang Tsai, Kun-Huang Chen, Chao-Ton Su, dan Hung-Chun Lin[11]. Pada penelitiannya Meng-Chang Tsai dkk. Mengusulkan metode baru dalam seleksi fitur, yaitu *Particle Swarm Optimization* dan *Algoritma C4.5* dengan menggunakan 5 dataset yang berbeda yang diperoleh dari *UCI machine learning databases*. Dengan menggunakan indeks akurasi untuk mengevaluasi efektivitas metode yang diusulkan. Dan hasil dari PSO dan algoritma C4.5 dibandingkan dengan *Regresi Logistitik (LR)*, *Back Propagation Neural Network (BPNN)*, *Support Vector Machine (SVM)*, dan *Decision Tree C4.5*. Kemudian, setelah melakukan eksperimen dari 5 jenis data set *BPNN* menghasilkan akurasi 75,38%, *LR* menghasilkan akurasi 83,59%, *SVM* menghasilkan akurasi 86,15%, *C4.5* menghasilkan akurasi 82,56%, dan *PSO+C4.5* menghasilkan akurasi 90,77%. Dari hasil akurasi yang dipaparkan, penulis pada penelitiannya menyimpulkan bahwa, *Particle Swarm Optimization (PSO)* + *Algoritma C4.5* masih melebihi metode yang lain dengan hasil akurasi mencapai 90,77%.

2.2. Landasan Teori

2.1.1 Data Mining

Fajar Astuti Hermawati [8] di dalam bukunya mengatakan, data mining adalah proses yang mempekerjakan satu atau lebih teknik pembelajaran computer (*machine learning*) untuk menganalisis dan mengekstraksi pengetahuan (*knowledge*) secara otomatis. Definisi lain diantaranya adalah pembelajaran berbasis induksi (*induction-based learning*) adalah proses pembentukan definisi-definisi konsep umum yang dilakukan dengan cara mengobservasi contoh-contoh spesifikasi dari konsep-konsep metode santifik pada data mining. Dalam konteks ini data mining merupakan satu langkah dari proses KDD.

2.1.2 Klasifikasi

Klasifikasi merupakan proses pembelajaran suatu fungsi (*target*) f yang memetakan tiap himpunan atribut x ke satu dari label kelas y yang didefinisikan sebelumnya. Fungsi target disebut dengan model klasifikasi.

Hermawati[8] di dalam bukunya mengatakan, bahwa teknik klasifikasi (*classiier*) adalah pendekatan

sistematis untuk membangun model klasifikasi dari suatu himpunan data masukan. Tiap teknik membutuhkan algoritma pembelajaran (*learning algorithm*) untuk mendapatkan suatu model yang paling memenuhi hubungan antara himpunan atribut dengan label kelas dalam data masukan. Tujuan dari algoritma pembelajaran adalah untuk membangun model yang secara umum berkemampuan baik, yaitu model yang dapat memprediksi label kelas dari *record* yang tidak diketahui kelas sebelumnya dengan lebih akurat.

2.1.3 Decision Tree

Pohon Keputusan (*Decision Tree*) merupakan metode klasifikasi dan prediksi yang sangat kuat dan terkenal. Metode pohon keputusan mengubah fakta yang sangat besar menjadi pohon keputusan yang merepresentasikan aturan. Aturan dapat dengan mudah dipahami dengan bahasa alami. Aturan ini juga dapat diekspresikan dalam bentuk bahasa basis data seperti SQL untuk mencari *record* pada kategori tertentu. Pohon keputusan juga berguna untuk mengeksplorasi data, menemukan hubungan tersembunyi antara sejumlah calon variabel input dengan sebuah variabel target. Karena pohon keputusan memadukan antara eksplorasi data dan pemodelan, pohon keputusan ini sangat bagus sebagai langkah awal dalam proses pemodelan bahkan ketika dijadikan sebagai model akhir dari beberapa teknik lain [9].

Untuk menghitung *entropy* diberikan rumus sebagai berikut :

$$1 \quad (1)$$

Dimana,

S = himpunan kasus

n = jumlah partisi S

p_i = proporsi dari S_i terhadap S

Untuk menghitung *Gain* diberikan rumus sebagai berikut:

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^n \frac{n_i}{n} \log_2 \left| \frac{p_i}{s} \right| * Entropy(S_i) \dots\dots\dots(2)$$

Dimana

S = himpunan kasus

A = Atribut

n = jumlah partisi

$|S_i|$ = jumlah kasus pada partisi ke- i

$|S|$ = jumlah kasus dalam S

Untuk menghitung *gain ratio* perlu diketahui terlebih dahulu suatu term baru yang disebut dengan *Split Information* dengan rumus :

$$SplitInfo(S, A) = - \sum_{i=1}^n \frac{S_i}{S} \log_2 \frac{S_i}{S} \dots\dots\dots(3)$$

Selanjutnya menghitung *Gain ratio* dengan rumus :

$$Gain Ratio(S, A) = \frac{Gain(S, A)}{SplitInformation(S, A)} \dots\dots\dots(4)$$

2.1.4 Particle Swarm Optimization (PSO)

Particle Swarm Optimization (PSO) merupakan teknik optimasi yang berbasis populasi yang di usulkan oleh Eberhart dan Kennedy pada tahun 1995. Metode ini terinspirasi pada perilaku sosial sekawan burung dan ikan [11]. Dalam mencari solusi yang optimal, partikel tersebut bergerak pada arah yang terbaik sebelumnya, posisi terbaik secara global. Seperti Ke- i dinyatakan $X_i = [X_{i1}, X_{i2}, \dots, X_{iD}]$ dalam ruang dimensi D . Posisi terbaik sebelumnya di simpan sebagai $pbest$ sebagai $pbest = [pbest_1, pbest_2, \dots, pbest_d]$. Indeks partikel terbaik dari semua di antara kawan group dinyatakan sebagai $gbest$. Kecepatan partikel dinyatakan sebagai $V_i = [V_{i1}, V_{i2}, \dots, V_{iD}]$. Modifikasi

kecepatan dan posisi partikel dapat di hitung menggunakan jarak $pbest$, $gbest$ seperti ditunjukkan persamaan berikut:

$$v_{i,d} = w * v_{i,d} + c1 * R * (pbest_{i,d} - x_{i,d}) + c2 * R * (gbest_d - x_{i,d}) \quad (5)$$

$$x_{i,d} = x_{i,d} + v_{i,d} \quad (6)$$

Keterangan:

- $V_{i, d}$ = Kecepatan partikel ke-i pada iterasi ke-i
- w = Faktor bobot inersia
- $c1, c2$ = Konstanta akselerasi (learning rate)
- R = Bilangan random (0-1)
- $X_{i, d}$ = Posisi saat ini dari partikel ke-i pada iterasi ke-i
- $pbest_i$ = Posisi terbaik sebelumnya dari partikel ke-i
- $gbest_i$ = Partikel terbaik di antara terbaik diantara semua partikel dalam satu kelompok atau populasi.

2.1.5 Confusion Matrix

Confusion Matrix adalah suatu metode yang biasanya digunakan untuk melakukan perhitungan akurasi pada konsep data mining. Rumus ini melakukan perhitungan dengan 3 keluaran, yaitu : *recall*, *precision*, *accuracy*.

Dalam model klasifikasi data mining, evaluasi yang digunakan untuk mengukur kinerja klasifikasi salah satunya adalah dengan menggunakan *precision* dan *recall*. *Precision* adalah cara pengukuran algoritma dalam mendapatkan data yang tepat [9]. Sementara *Recall* adalah cara pengukuran algoritma dalam mendapatkan data tepat yang terbaca.

Persamaan *Precision* adalah dengan menjumlah data yang true positive dibagi dengan jumlah data yang dikenali sebagai positif. *recall* adalah jumlah data yang true positive dibagi dengan jumlah data yang sebenarnya positif

Untuk menghitung Precision digambarkan dalam persamaan berikut :

$$Precision = \frac{TP}{TP+FP} \dots\dots\dots(2.7)$$

Sementara menghitung Recall di gunakan persamaan berikut :

$$Recall = \frac{TP}{TP+FN} \dots\dots\dots(2.8)$$

Kemudian untuk mengetahui tingkat akurasi klasifikasi dapat dihitung dengan persamaan berikut :

$$Accuracy = \frac{TP+TN}{TP+FP+TN+ FN} \dots\dots\dots(2.9)$$

Accuracy adalah prosentase dari total prediksi yang benar diidentifikasi.

Tabel 1. Actual Class dan Predicted Class

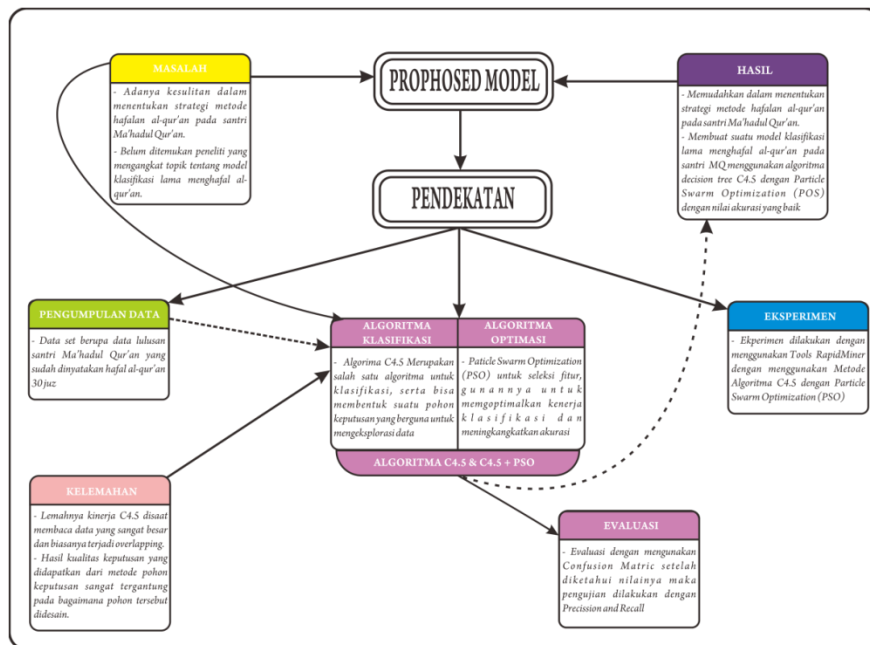
	Actual Class		
		1	0
Predicted Class	1	TP	FP
	0	FN	TN

Dari tabel tersebut diatas terdapat beberapa angka yang dapat dianalisa. Tabel terdiri dari dua buah

data yaitu data kelas yang dihasilkan dari *classifier* (*Predicted Class*) dan data kelas asli yang telah diketahui (*Actual Class*). Jika kelas yang dihasilkan classifier sama dengan actual kelas dari data maka data termasuk dalam TP (True Positive) untuk kelas positive (1) atau TN (True Negative) untuk kelas negative (0).

2.3. Kerangka Pemikiran Model Klasifikasi Algoritma C4.5 dengan Particle Swarm Optimization

Pada penelitian ini, dataset yang digunakan adalah dataset lulusan santri Ma'hadul Qur'an yang telah berhasil menghafalkan al-qur'an 30 juz. Dalam melakukan penelitian ini, dibuat sebuah kerangka pemikiran yang berguna sebagai pedoman dalam melakukan penelitian sehingga penelitian yang dilakukan dapat berjalan konsisten. Berikut adalah kerangka pemikiran dalam penelitian ini.



Gambar1. Kerangka Pemikiran

3. METODE PENELITIAN

3.1. Pengumpulan Data

Dalam penelitian ini, menggunakan data lulusan santri yang berada di Asrama Ma'hadul Qur'an PP Salafiyah Syafi'iyah Sukorejo. Penelitian ini, membutuhkan *data primer* yang didapat dari bagian *Kaur. Kesiswaan MQ*. Dataset yang diperoleh dari tahun 2008 sampai tahun 2015 berjumlah 1644 data lulusan santri yang dinyatakan lulus dari asrama Ma'hadul Qur'an.

3.2. Eksperimen

Eksperimen dilakukan melalui tahapan training dan testing. Namun dari jumlah data itu yang dijadikan sebagai data training berjumlah 200 data, sedang data testing menggunakan 20 data. Dan sudah bisa dipastikan data sudah benar-benar bisa sudah bisa diolah dengan baik dengan atribut-atribut yang sudah ditentukan. Algoritma yang digunakan adalah *Decision Tree C4.5* dan *Algoritma Particle Swarm Optimization (PSO)*.

3.3. Evaluasi

Pengujian model akan dilakukan proses perbandingan dalam *Algoritma Decision Tree C4.5* dengan *Algoritma Particle Swarm Optimization (PSO)* ke model pembelajaran. Evaluasi model dilakukan dengan menggunakan *confusion matrix dan precisin and recall*.

Evaluasi model dengan dilakukan penghitungan fungsi *precision* dan *recall* terhadap model yang dibangun. Dengan membandingkan hasil prediksi positif benar dan hasil positif salah pada suatu kelas klasifikasi.

Misal untuk menghitung Precision dari kelas cepat digambarkan dalam persamaan berikut:

$$Precision_{cepat} = \frac{TP}{TP+FP}$$

Sementara menghitung Recall dari kelas cepat dapat di gunakan persamaan berikut :

$$Recall_{cepat} = \frac{TP}{TP+FN}$$

4. HASIL DAN PEMBAHASAN

Kinerja algoritma C4.5 dalam fase pengujian ini bertujuan untuk mengetahui hasil dari mengklasifikasi data lamanya menghafal Al-qur'an Santri Ma'hadul Qur'an kedalam kelas yang ditentukan. Dengan melakukan dua tahap pengujian yang bertujuan untuk mengetahui akurasi dengan menggunakan metode algoritma C4.5 dan komparasi algoritma C4.5 dengan *Particle Swam Optimization (PSO)* dengan data traning menggunakan 200 dataset dan data testing sebanyak 20 dataset.

4.1. Pengujian Algoritma C4.5

Langkah pertama dalam pengujian algoritma *Decision Tree C4.5* adalah dengan cara menghitung nilai *entropy* dari semua class dan menghitung class. Kemudian dilakukan perhitungan *gain* untuk di setiap atribut, karena nilai *gain* tertinggi yang akan menjadi akar. Dengan menggunakan persamaan 1 maka menghasilkan nilai *entropy* total sebagai berikut :

$$Entropy(\text{Total}) = \left(-\frac{119}{200} \times \log_2 \left(\frac{119}{200} \right) \right) + \left(-\frac{81}{200} \times \log_2 \left(\frac{81}{200} \right) \right)$$

$$Entropy(\text{Total}) = 0,9738$$

Untuk mengetahui sebuah nilai *gain* atribut membutuhkan nilai *entropy* dari setiap kasus pada atribut tersebut. Maka, *entropy* harus diketahui terlebih dahulu dengan menggunakan *persamaan 1*.

Langkah kedua yaitu dengan menghitung nilai *gain* dari setiap atribut dengan menggunakan persamaan 2, sehingga menghasilkan nilai sebagai berikut :

$$Gain(\text{Total, Tempat Tinggal}) = 0,9738 - \left(\left(\frac{133}{200} \times 0,9434 \right) + \left(\frac{67}{200} \times 0,9998 \right) \right)$$

$$Gain(\text{Total, Tempat Tingal}) = 1,9361$$

$$Gain(\text{Total, Uang Saku}) = 0,9738 - \left(\left(\frac{64}{200} \times 0,7856 \right) + \left(\frac{101}{200} \times 0,9623 \right) + \left(\frac{35}{200} \times 0,5127 \right) \right)$$

$$Gain(\text{Total, Uang Saku}) = 1,8008$$

$$Gain(\text{Total, Usia}) = 0,9738 - \left(\left(\frac{50}{200} \times 0,9988 \right) + \left(\frac{55}{200} \times 0,7568 \right) + \left(\frac{95}{200} \times 0,9980 \right) \right)$$

$$Gain(\text{Total, Usia}) = 1,9057$$

$$Gain(\text{Total, Pendidikan}) = 0,9738 -$$

$$\left(\left(\frac{37}{200} \times 0,9353 \right) + \left(\frac{32}{200} \times 0,9887 \right) + \left(\frac{18}{200} \times 0,9183 \right) + \left(\frac{39}{200} \times 0,8905 \right) + \left(\frac{33}{200} \times 0,3298 \right) + \left(\frac{41}{200} \times 0,9012 \right) \right)$$

$$Gain(\text{Total, Pendidikan}) = 1,8005$$

$$\text{Gain (Total, Nilai Tes)} = 0,9738 - \left(\left(\frac{52}{200} \times 0,9118 \right) + \left(\frac{92}{200} \times 0,9915 \right) + \left(\frac{56}{200} \times 0,9769 \right) \right)$$

$$\text{Gain (Total, Nilai Tes)} = 1,9405$$

Perhitungan *node akar* menunjukkan bahwa *gain* tertinggi ada pada fitur “*Nilai Tes*” dengan nilai *gain* sebesar 1,9405. Sehingga fitur “*Nilai Tes*” dijadikan sebagai *node akar*. Untung selanjutnya, dihitung posisi *split* untuk fitur tersebut dengan menggunakan *persamaan 3* dan menghitung *ratio gain* dengan menggunakan *persamaan 4*. hasil penghitungan *split* untuk fitur “*Nilai Tes*” adalah sebagai berikut :

$$\text{SplitInfo}(\text{Semua, Nilai Tes}) = - \left(\left(\frac{52}{200} \right) \times \text{Log}_2 \left(\frac{52}{200} \right) \right) + \left(\left(\frac{92}{200} \right) \times \text{Log}_2 \left(\frac{92}{200} \right) \right) + \left(\left(\frac{56}{200} \right) \times \text{Log}_2 \left(\frac{56}{200} \right) \right)$$

$$\text{SplitInfo}(\text{Semua, Nilai Tes}) = 1,5348$$

$$\text{Gain Ratio (Semua, Nilai Tes)} = \frac{\text{Gain (Semua, Nilai Tes)}}{\text{Split Info (Semua, Nilai Tes)}}$$

$$\text{Gain Ratio (Semua, Nilai Tes)} = \frac{1,9405}{1,5348}$$

$$= 1,2643$$

Selanjutnya dengan cara yang sama dilakukan penghitungan untuk node berikutnya sampai proses pembentukan tree selesai, pada penelitian ini penghitungan dilakukan sampai dengan node 60.

Selanjutnya dilakukan evaluasi dengan confusion matrik. Confusion matrix adalah suatu metode yang biasanya digunakan untuk melakukan perhitungan akurasi pada konsep data mining. Rumus ini melakukan perhitungan dengan 3 keluaran, yaitu: recall, precision, accuracy. Untuk penghitungan confusion matrik dari data yang ada diambil 20 data testing, dan hasil pencocokan dengan hasil model C4.5. sehingga diperoleh penghitungan *confusion matrix* sebagai berikut :

Tabel 2. Model *confusion matrix* pada pengujian Algoritma C4.5

Predicated Class	Actual Class	
	Cepat	Lambat
Cepat	12	1
Lambat	3	4

pada tabel 4.13 dapat diketahui untuk kelas cepat terdapat 12 data terprediksi secara benar di kelasnya, dari 20 data testing. Kemudian, evaluasi klasifikasi Algoritma C4.5 dapat dihitung dengan persamaan *Precision and Recall* berikut ini.

$$\text{Precision} = \frac{12}{12+1} \times 100\% = 92,3\%$$

$$\text{Recall} = \frac{12}{12+3} \times 100\% = 80\%$$

$$\text{Accuracy} = \frac{12+4}{20} \times 100\% = 80\%$$

4.2. Pengujian Algoritma Particle Swarm Optimization (PSO)

Selanjutnya langkah-langkah untuk menghitung pada metode *Particle Swarm Optimization (PSO)* dengan tujuan untuk mencari solusi yang lebih optimal adapun langkah pertama yaitu menghitung nilai fitness dengan menggunakan satu variabel dengan formula sebagai berikut: $\min f(x) = (1 - x)^2$

Untuk menentukan populasi awal berdasarkan jumlah $n=6$ sebagai berikut :

$$\begin{aligned} x_1(0) &= 0,45 \\ x_2(0) &= 0,55 \\ x_3(0) &= 0,70 \\ x_4(0) &= 0,80 \\ x_5(0) &= 0,90 \\ x_6(0) &= 0,95 \end{aligned}$$

Selanjutnya melakukan perhitungan dengan nilai fitness pada setiap partikel $X_j(0) = j=1,2,3,\dots,6$

$$\begin{aligned} f_1(0) &= f(0,45) = (1-0,45)^2 = 0,30 \\ f_2(0) &= f(0,55) = (1-0,55)^2 = 0,20 \\ f_3(0) &= f(0,70) = (1-0,70)^2 = 0,09 \\ f_4(0) &= f(0,80) = (1-0,80)^2 = 0,04 \\ f_5(0) &= f(0,90) = (1-0,90)^2 = 0,01 \\ f_6(0) &= f(0,95) = (1-0,95)^2 = 0,00 \end{aligned}$$

Selanjut menentukan kecepatan awal $v_1(0), v_2(0), \dots, v_7(0) = 0$ Iterasi 1 : Mencari nilai $\sim\sim Best_1(0), pBest_2(0), \dots, pBest_7(0)$ dan $gBest$.

$$\begin{aligned} pBest_1(1) &= 0,45 & gBest(1) &= 0,95 \\ pBest_2(1) &= 0,55 \\ pBest_3(1) &= 0,70 \\ pBest_4(1) &= 0,80 \\ pBest_5(1) &= 0,90 \\ pBest_6(1) &= 0,95 \end{aligned}$$

Kemudian melakukan perhitungan nilai kecepatan partikel $v_j(1)$ dengan $c_1 = 1, c_2 = 1$ nilai random $r_1 = 0,2, c_2 = 0,5$, dan nilai inertia 1 dengan formula 5 maka akan dihasilkan perhitungan dibawah ini :

$$\begin{aligned} V_1(1) &= 0 + 5 \times 0,2 \times (0,45+0,45) + 1 \times 0,5 (0,95-0,45) = 0,25 \\ V_2(1) &= 0 + 5 \times 0,2 \times (0,55+0,55) + 1 \times 0,5 (0,95-0,55) = 0,20 \\ V_3(1) &= 0 + 5 \times 0,2 \times (0,70+0,70) + 1 \times 0,5 (0,95-0,70) = 0,13 \\ V_4(1) &= 0 + 5 \times 0,2 \times (0,80+0,80) + 1 \times 0,5 (0,95-0,80) = 0,08 \\ V_5(1) &= 0 + 5 \times 0,2 \times (0,90+0,90) + 1 \times 0,5 (0,95-0,90) = 0,03 \\ V_6(1) &= 0 + 5 \times 0,2 \times (0,95+0,95) + 1 \times 0,5 (0,95-0,95) = 0,00 \end{aligned}$$

Langkah selanjutnya adalah melakukan perhitungan pada posisi baru pada masing-masing partikel berdasarkan hasil nilai kecepatan dengan formula 6

$$\begin{aligned} x_1(1) &= 0,45 + 0,25 = 0,70 \\ x_2(1) &= 0,55 + 0,20 = 0,75 \\ x_3(1) &= 0,70 + 0,13 = 0,83 \\ x_4(1) &= 0,80 + 0,08 = 0,88 \\ x_5(1) &= 0,90 + 0,03 = 0,93 \end{aligned}$$

$$x_6(1) = 0,95 + 0,00 = 1,90$$

Selanjutnya untuk mengevaluasi pada nilai fungsi dengan melakukan perhitungan nilai fitness untuk setiap partikel $X_j(0)=j=1,2,3,\dots,6$

$$f_1(0) = f(0,70) = (1-0,70)^2 = 0,09$$

$$f_2(0) = f(0,75) = (1-0,75)^2 = 0,06$$

$$f_3(0) = f(0,83) = (1-0,83)^2 = 0,03$$

$$f_4(0) = f(0,88) = (1-0,88)^2 = 0,02$$

$$f_5(0) = f(0,93) = (1-0,93)^2 = 0,01$$

$$f_6(0) = f(1,90) = (1-1,90)^2 = 0,81$$

Pada hasil perhitungan diatas, bahwa nilai fungsi sebelumnya $f(0)$ dibandingkan dengan nilai fungsi $f(1)$, maka dapat dirumuskan nilai fungsi $f(1)$ tidak menghasilkan nilai terbaik dari iterasi sebelumnya $f(0)$ dan dari hasil perhitungan dapat disimpulkan apakah nilai x sudah konvergen atau memusat dimana x memiliki nilai yang saling berdekatan dengan lainnya. Apabila nilai x masih belum tercapai konvergen, maka perlu dilakukan iterasi selanjutnya, apabila nilai x sudah tercapai konvergen, maka dari partikel x digunakan untuk melakukan penyeleksian dan hasil dari seleksi tersebut akan digunakan sebagai atribut pada training C4.5 berbasis *Particle Swarm Optimization* (PSO).

Tabel 3. Hasil percobaan C4.5 dengan *Particle Swarm Optimization*

NO	Pop Size	Max Mum Of Generatin	Inertia Weight	Local Best	Global Best	Akurasi	Proses Waktu	Atribut Terpilih
1	6	30	1	0,45	0,95	70.0	12	2
2	6	30	1	0,55	0,95	70.0	11	2
3	6	30	1	0,70	0,95	70.0	11	5
4	6	30	1	0,80	0,95	70.0	10	2
5	6	30	1	0,90	0,95	80.0	9	0
6	6	40	1	0,95	0,95	80.0	15	3
7	6	40	1	0,45	0,95	75.0	15	3
8	6	40	1	0,55	0,95	80.0	15	5
9	6	40	1	0,70	0,95	80.0	14	3
10	6	40	1	0,80	0,95	85.0	12	0
11	6	50	1	0,90	0,95	75.0	19	3
12	6	50	1	0,95	0,95	80.0	17	3
13	6	50	1	0,45	0,95	85.0	19	4
14	6	50	1	0,55	0,95	86.0	17	3
15	6	50	1	0,70	0,95	87.0	15	0

Dari table 3. percobaan dilakukan selama 15 kali karena setiap percobaan mendapatkan nilai akurasi dan tingkat kesalahan yang berbeda yang disebabkan oleh data yang diambil secara acak pada setiap percobaan walaupun tetap dibagi kedalam 15 bagian yang hampir sama jumlahnya dan diambil nilai akurasi yang tertinggi. Adapun hasil percobaan C4.5 berbasis *Particle Swarm Optimization* dengan maximum number of generation 30 dengan jumlah atribut sebanyak 6, $w=1$, $c1=3$, $c2=1$ adalah yang paling maksimal, yaitu : Accuracy 87 %, Recall 92,33 %, Precision 65,33 %, AUC 0,867.

4.3. Perbandingan Decision Tree C4.5 dan Decision Tree C4.5 berbasis Particle Swarm Optimization

Pada penelitian klasifikasi Algoritma C4.5 dengan Particle Swarm Optimization untuk Klasifikasi Lama Menghafal Alqur'an pada Santri Ma'hadul Qur'an, menghasilkan perbandingan sebagai berikut :

Tabel 4. Perbandingan Hasil C4.5 Dan C4.5/PSO

	Accuracy	AUC
<i>Decision Tree</i>	80%	0,817
<i>PSO + Particel Swarm Optimization</i>	87%	0,867

Dari table model yang dihasilkan diuji untuk mendapatkan nilai *accuracy* dan AUC dari setiap algoritma, sehingga didapat pengujian dengan menggunakan *C4.5* didapat nilai *accuracy* adalah 80,0% dengan nilai AUC 0,817 Sedangkan pengujian dengan menggunakan *C4.5* berbasis *Particle Swarm Optimization* didapatkan nilai *accuracy* 87,0% dengan nilai AUC 0,867.

5. PENUTUP

5.1. Kesimpulan

Pada penelitian klasifikasi Algoritma C4.5 dengan Particle Swarm Optimization untuk Klasifikasi Lama Menghafal Al-qur'an pada Santri Ma'hadul Qur'an, dengan menggunakan 200 data training dan 20 data testing setelah dilakukan pengujian dengan menggunakan algoritma C4.5 menghasilkan *Accuracy* 80 % dengan nilai *AUC* 0,817. Sedangkan pengujian dengan menggunakan algoritma C4.5 dan Particle Swarm Optimization menghasilkan *Accuracy* 87 % dengan nilai *AUC* 0,867. Dari hasil penghitungan tersebut, dapat ditarik kesimpulan bahwa pengujian klasifikasi lama menghafal Alqur'an pada santri Ma'hadul Qur'an menggunakan Algoritma C4.5 dengan Particle Swarm Optimization didapat bahwa pengujian C4.5 dengan Particle Swarm Optimization lebih baik dari pada C4.5 itu sendiri.

5.2. Saran

Sebuah algoritma data mining, dalam pengukuran kinerja bisa diimplementasikan berdasarkan beberapa kriteria, diantaranya keakuratan prediksi, kecepatan/efisiensi, kehandalan, skalabilitas dan interpretabilitas [4]. Sedangkan dalam penelitian ini menggunakan satu kriteria saja, yaitu berdasarkan akurasi. Oleh sebab itu, penelitian lain bisa menggunakan kriteria yang lainnya.

UCAPAN TERIMA KASIH

Penulis menyampaikan terima kasih kepada Rektor Universitas Dian Nuswantoro, Kaprodi MTI Universitas Dian Nuswantoro, dan semua pihak yang telah membantu penulis dalam penyusunan Tesis ini.

PERNYATAAN ORISINALITAS

“ Saya menyatakan dan bertanggung jawab dengan sebenarnya bahwa artikel ini adalah hasil karya saya sendiri kecuali cuplikan dan ringkasan yang masing-masing telah saya jelaskan sumbernya”

[FIRMAN SANTOSO]

DAFTAR PUSTAKA

- [1] IAbadSukorejo. (2013, April) PP. Salafiyah Syafi'iyah Sukorejo. [Online]. [HYPERLINK "http://sukorejo.com/pondok-pesantren-salafiyah-syafiiyah-asembagus-situbondo-jatim/"](http://sukorejo.com/pondok-pesantren-salafiyah-syafiiyah-asembagus-situbondo-jatim/)

<http://sukorejo.com/pondok-pesantren-salafiyah-syafiiyah-asembagus-situbondo-jatim/>

- [2] Drs KHM Hasan Basri Lc, *KHR. As'ad Syamsul Arifin Riwayat Hidup dan Perjuangannya*, 1st ed., Drs Chairul Anam, Ed. Situbondo, Indonesia: PONDOK PESANTREN SALAFIYAH SYAFIYAH, 1994.
- [3] Fithriani Gade, "IMPLEMENTASI METODE TAKRĀR DALAM PEMBELAJARAN MENGHAFAL AL-QUR'AN," *Jurnal Ilmiah*, vol. XIV, no. 2, pp. 4-5, Februari 2014.
- [4] Sukardi, Abd Syukur, and Catur Supriyanto, "KLASIFIKASI SPAM EMAIL MENGGUNAKAN ALGORITMA C4.5 DENGAN SELEKSI FITUR," *Jurnal Teknologi Informasi*, vol. X, no. 1, April 2014.
- [5] Liliana Swastina, "Penerapan Algoritma C4.5 Untuk Penentuan Jurusan Mahasiswa," *Jurnal GEMA AKTUALITA*, vol. II, no. 1, Juni 2013.
- [6] Zhiwu Liu and Xiuzhi Zhang, "Prediction and Analysis for Students' Marks Based on Decision Tree Algorithm," *2010 Third International Conference on Intelligent Networks and Intelligent Systems*, pp. 339-341, 2010.
- [7] Daniel T. Larose, *DISCOVERING KNOWLEDGE IN DATA An Introduction to Data Mining*. United States of America: Willey-Interscience, 2005.
- [8] Fajar Astuti Hermawati, *Data Mining*, Putri Christian, Ed. Yogyakarta, Indonesia: CV ANDI OFFSET, 2013.
- [9] Eko Prasetyo, *Data Mining Mengolah Data Menjadi Informasi menggunakan Matlab*, 2nd ed., Aldo Sahala, Ed. Yogyakarta, Indonesia: CV Andi Offset, 2014.
- [10] S. B. Kotsiantis, "Supervised Machine Learning: A Review of Classification Techniques," *Department of Computer Science and Technology*, pp. 249-268, Juli 2007.
- [11] Meng-Chang Tsai, Kun-Huang Chen, Chao-Ton Su, and Hung-Chun Lin, "An Application of PSO Algorithm and Decision Tree for Medical Problem," *2nd International Conference on Intelligent Computational Systems*, pp. 124-126, Oktober 2012.