

## Penerapan Data Mining Untuk Klasifikasi Gaya Belajar Siswa Menggunakan Algoritma C4.5

### *Application Of Data Mining For Student Learning Style Classification Using C4.5 Algorithm*

Dinda Novita Sari<sup>1</sup>, Hardian Oktavianto<sup>2\*</sup>, Ilham Saifudin<sup>3</sup>

<sup>1</sup>Mahasiswa Fakultas Teknik, Universitas Muhammadiyah Jember

Email: [dindanovitasari1710@gmail.com](mailto:dindanovitasari1710@gmail.com)

<sup>2</sup>Dosen Fakultas Teknik, Universitas Muhammadiyah Jember\* Koresponden Author

Email : [Hardian@unmuhjember.ac.id](mailto:Hardian@unmuhjember.ac.id)

<sup>3</sup>Dosen Fakultas Teknik, Universitas Muhammadiyah Jember

Email : [ilham.saifudin@unmuhjember.ac.id](mailto:ilham.saifudin@unmuhjember.ac.id)

#### ABSTRAK

Data mining merupakan serangkaian proses yang digunakan untuk mendapatkan informasi yang berguna dari database yang besar. SMP Negeri 2 Mayang merupakan Sekolah Menengah Pertama yang ada di kecamatan Mayang Kabupaten Jember. Pada proses pembelajaran di SMP Negeri 2 Mayang, masih banyak murid yang tidak fokus atau tidak memiliki keinginan untuk belajar. Kemungkinan besar ketidakcocokan gaya belajar dan metode belajar yang guru berikan. Oleh karena itu, penulis melakukan analisis data mining terhadap data murid agar data tersebut dapat berubah menjadi informasi yang berharga bagi murid dan guru di SMP Negeri 2 Mayang. Penulis menggunakan hasil data kuisioner pada siswa-siswi kelas 8 & 9 sebanyak 138 data siswa-siswi. Dalam melakukan analisis ini, penulis menggunakan *tools Weka* dengan algoritma C4.5 dengan 6 atribut. Perhitungan menggunakan metode *cross validation* menghasilkan rata-rata nilai *Correctly Classified Instances* 80,03% dan *Incorrectly Classified Instances* 19,97%. Berdasarkan hasil evaluasi menggunakan algoritma C4.5 gaya belajar pada siswa berpengaruh dengan cara mengingat dan hobi pada masing-masing siswa.

**Kata kunci** : Data mining, Algoritma C4.5, Gaya Belajar

#### ABSTRACT

*Data mining is a series of processes used to obtain useful information from large databases. SMP Negeri 2 Mayang is a junior high school in the Mayang sub-district, Jember Regency. In the learning process at SMP Negeri 2 Mayang, there are still many students who are not focused or have no desire to learn. Most likely the incompatibility of learning styles and learning methods that the teacher provides. Therefore, the authors conduct data mining analysis of student data so that the data can be turned into valuable information for students and teachers at SMP Negeri 2 Mayang. The author uses the results of questionnaire data on 8th & 9th grade students as many as 138 students' data. In conducting this analysis, the author uses the Weka tool with the C4.5 algorithm with 6 attributes. Calculations using the cross validation method produce an average value of 80.03% Correctly Classified Instances and 19.97% Incorrectly Classified Instances. Based on the results of the evaluation using the C4.5 algorithm, the students' learning styles have an effect on how to remember and hobbies on each student.*

*Keywords: Data mining, C4.5 Algorithm, Learning Style*

## 1. PENDAHULUAN

Belajar ialah suatu proses untuk memperoleh ilmu pengetahuan. Proses ini selalu terjadi didalam kehidupan sehari-hari. Belajar dilakukan oleh semua orang terutama siswa sebagai peserta didik dilingkungan Pendidikan. Kebanyakan siswa mengalami kesulitan dalam belajar, ada yang memiliki ingatan yang buruk, bermasalah dengan konsentrasi yang sehingga penyerapan informasi siswa berkurang.

Tipe – tipe gaya belajar cenderung berbeda satu sama lain. Fakta tersebut kemudian menjadikan gaya belajar salah satu acuan bagi para guru dalam menentukan metode pembelajaran yang sekiranya cocok diterapkan di kelas. Alangkah tidak bijaknya jika seorang guru hanya menggunakan satu metode mengajar saja secara monoton dalam setiap mengajarnya. Dengan kata lain, guru tersebut hanya mengakomodasikan salah satu dari sekian banyak gaya belajar siswa.

Secara umum, bahwa gaya belajar dikelompokkan menjadi 3 yaitu gaya belajar visual, gaya belajar kinestetik dan gaya belajar auditori. Gaya belajar visual yaitu proses belajar seperti mengamati, menggambar serta melihat. Gaya belajar kinestetik adalah gaya belajar bergerak, bekerja serta menyentuh. Sedangkan gaya belajar auditori yaitu gaya belajar yang lebih suka menyimak. Gaya belajar ini lebih gemar membaca lantang serta mendengarkan sesuatu, cara berbicara yang lancar dan lebih suka dengan musik. (Prasetyo et al., 2021)

SMP Negeri 2 Mayang adalah salah satu sekolah negeri yang ada di kota Jember. Dalam metode pembelajaran disetiap guru pasti berbeda. Guru ada yang mengajar dengan cara ceramah atau bahkan memanfaatkan perangkat yang ada telah disediakan. Guru yang mengajar dengan metode ceramah sebesar 80% dan untuk guru yang mengajar menggunakan perangkat multimedia sebesar 20%. Permasalahan secara umum yang ada di SMP Negeri 2 Mayang khususnya adalah gaya belajar siswa yang berbeda-beda. Dari hasil observasi masih banyak siswa siswa yang bermain – main maupun tidak berminat disaat mengikuti proses belajar mengajar.

Berdasarkan masalah tersebut, penulis mencoba melakukan penelitian untuk

menganalisa algoritma C4.5. Adapun sumber dataset adalah data hasil kuisioner siswa – siswi SMP Negeri 2 Mayang. Perhitungan akurasi, recall dan presisi yang diuji akan menghasilkan informasi yang efisien, akurat. Atas dasar inilah, penulis tertarik untuk mengambil judul tugas akhir :“Penerapan Data Mining Untuk Memprediksi Gaya Belajar Siswa Menggunakan Algoritma C4.5 (Studi Kasus : SMP Negeri 2 Mayang)”.

Rumusan Masalah dari penelitian ini adalah sebagai berikut:

1. Berapa accuracy, Precision dan Recall dalam memprediksi gaya belajar siswa kelas 8 dan 9 di SMP Negeri 2 Mayang?
2. Apa saja yang mempengaruhi siswa dengan gaya belajar visual, kinestetik dan auditori?

Tujuan Penelitian ini adalah sebagai berikut :

1. Menghitung accuracy, precision dan recall dalam memprediksi gaya belajar siswa kelas 8 dan 9 di SMP Negeri 2 Mayang.
2. Untuk mengetahui pengaruh apa saja yang membuat gaya belajar siswa berbeda-beda.

Dalam penelitian ini terdapat beberapa Batasan masalah agar tidak menyimpang dari data di atas, berikut batasan masalah pada penelitian ini:

1. Penelitian dilakukan di SMP Negeri 2 Mayang.
2. Dataset yang digunakan data hasil kuisioner siswa SMP Negeri 2 Mayang yang datanya sebanyak 138 data.
3. Atribut yang digunakan pada penelitian ini diantaranya yaitu metode belajar, cara membaca, cara mengingat, cara berbicara, cara memahami, hobi dan gaya belajar.
4. *Tools* yang digunakan adalah Weka.
5. Pengujian dilakukan menggunakan *cross Validation*.

## 2. TINJAUAN PUSTAKA

### a. Data Mining

Menurut Gartner Group data mining adalah suatu proses menciptakan ikatan yang berarti, pola serta kecenderungan dengan mengecek dalam sekumpulan besar informasi yang tersimpan dalam penyimpanan, dengan memakai teknik pengenalan pola semacam

metode statistik serta matematika (Informasi & Informatika, 2005).

#### b. Decision Tree

Klasifikasi adalah proses menemukan model (fungsi) yang menjelaskan dan membedakan kelas-kelas atau konsep, dengan tujuan agar model yang diperoleh dapat digunakan untuk memprediksi kelas atau objek yang memiliki label kelas tidak diketahui. Model yang turunan didasarkan pada analisis dari training data (yaitu objek data yang memiliki label kelas yang diketahui) (Atmaja et al., 2018).

#### c. Pohon Keputusan

Pohon keputusan merupakan metode klasifikasi yang paling populer digunakan. Selain karena pembangunannya relatif cepat, hasil dari model yang dibangun mudah untuk dipahami. Pohon keputusan yang dapat disimpulkan aturan-aturan klasifikasi tertentu, salah satunya adalah algoritma C4.5 (Azwanti, 2018).

#### d. Algoritma C4.5

Algoritma C4.5 merupakan pengembangan dari ID3, dimana pengembangan dilakukan dalam hal, bisa mengatasi *missing* data, bisa mengatasi data kontinu dan pruning (Elisa, 2017).

Secara umum algoritma C4.5 untuk membangun pohon keputusan adalah sebagai berikut (Lukhayu Pritalia, 2018):

1. Menghitung Entropy total dari dataset dilanjutkan dengan entropy masing-masing atribut.
2. Setelah diperoleh *entropy* masing-masing atribut, menghitung *information gain* masing-masing.
3. Memilih atribut memiliki *information gain* paling besar sebagai akar.
4. Mengulangi perhitungan *entropy* dan *gain* untuk menentukan atribut berikutnya sebagai daun.

Untuk memilih atribut sebagai root berdasarkan nilai perolehan tertinggi dari atribut yang ada. Untuk menghitung profit digunakan rumus seperti di bawah ini. (Prayoga, 2018)

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

Keterangan:

Gain: Proses perhitungan untuk mendapatkan nilai Entropy yang terbesar

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 mendapatkan nilai Gain, Anda mencari nilai Entropy. Entropi digunakan untuk menentukan setiap atribut masukan informatif untuk menghasilkan suatu atribut. Rumus dasar Entropi adalah sebagai berikut (Rani, 2016):

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

Keterangan:

Entropy: Proses perhitungan untuk mengukur ketidakpastian variable acak

S: Himpunan kasus

n: Jumlah partisi S

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

#### e. Weka

Weka adalah sebuah paket *tools machine learning* praktis. Weka mampu menyelesaikan masalah – masalah data mining di dunia nyata, khususnya klasifikasi yang mendasari pendekatan-pendekatan machine learning. Perangkat lunak ini ditulis dengan hirarki class java dengan metode berorientasi objek dan dapat berjalan hampir semua platform (Nugraha et al., 2016).

#### f. Gaya Belajar

Gaya belajar adalah cara yang konsisten yang dilakukan oleh seorang peserta didik dalam menangkap stimulus / informasi, cara mengingat, berfikir dan memecahkan masalah (Kadir et al., 2020).

#### g. Cross Validation

*Cross Validation* merupakan salah satu teknik untuk menilai/memvalidasi keakuratan sebuah model yang dibangun berdasarkan dataset tertentu. Pembuatan model biasanya bertujuan untuk melakukan prediksi maupun klasifikasi terhadap suatu data baru yang boleh jadi belum pernah muncul di dalam dataset (Novianti & Santosa, 2016).

### h. Confusion Matrix

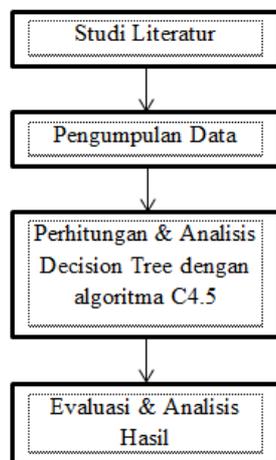
*Confusion matrix* merepresentasikan dan kondisi sebenarnya dari data yang dihasilkan oleh algoritma *machine learning*. Berdasarkan *confusion matrix* bisa menentukan akurasi, presisi, dan recall (Ginanjar, 2019).

$$\text{akurasi} = \frac{TP+TN}{TP+FP+FN+TN}$$

$$\text{Precision} = \frac{TP}{TP+FP}$$

$$\text{Recall} = \frac{TP}{TP+FN}$$

## 3. METODOLOGI



**Gambar 1.** Alur Penelitian

Sumber : Hasil Gambar menurut alur penelitian sendiri.

### 1. Studi Literatur

Tujuan dilakukan studi literatur untuk mendapatkan landasan – landasan yang digunakan untuk penelitian ini, studi literatur yang dilakukan adalah dengan mencari buku dan jurnal yang berhubungan dengan penelitian klasifikasi algoritma C4.5.

### 2. Pengumpulan data

Pengumpulan data yang dilakukan dalam penelitian ini diperoleh dari SMP Negeri 2 Mayang. Pengumpulan data dilakukan dengan wawancara langsung dengan Bapak Mahrus Varisi pada tanggal 15 September 2021 di SMP

Negeri 2 Mayang. Dataset berupa hasil kuisioner kepada siswa – siswi di SMP Negeri 2 Mayang yang terdiri dari 138 data siswa.

### 3. Menghitung *decision tree* dengan algoritma C4.5

Persiapan awal ditentukan atribut yang digunakan kemudian melakukan uji atribut yang digunakan, kemudian melakukan uji atribut dengan mencari nilai *gain* tertinggi berdasarkan perhitungan *entropy* dari masing-masing atribut. Apabila *gain* tertinggi maka *gain* tersebut akan menjadi *root* awal. Selanjutnya, dilakukan penentuan cabang dengan cara yang sama dengan melihat *gain* tertinggi dari tiap hasil partisi.

## 4. HASIL DAN PEMBAHASAN

Hasil pengujian adalah hasil akhir yang berarti penelitian telah selesai dilakukan, semua yang berkaitan dengan pengujian akan dibahas. Mulai dari pengolahan *Dataset* hingga pengujian data pada *Software*.

### a. Analisis Data

Dari data hasil kuisioner diatas proses yang perlu dilakukan selanjutnya itu pembersihan data yaitu dengan proses pembuangan noise dan data yang tidak konsisten atau tidak relevan. Yang dilakukan dalam proses ini yaitu dengan menggantikan beberapa kata yang salah ketik atau bahkan data yang tidak valid. Tahap selanjutnya yaitu data yang sudah terkumpulkan akan dilakukan seleksi data yaitu dengan mengambil beberapa atribut yang telah di validasikan oleh guru bimbingan konseling dan untuk atribut yang tidak di validasikan untuk proses klasifikasi maka atribut tersebut tidak perlu diujikan dengan algoritma C4.5. Dari 18 pertanyaan yang digunakan hanya 6 atribut itu metode belajar, cara membaca, cara mengingat, cara berbicara, cara memahami dan hobi. Dari proses transformasi data yaitu melakukan pengubahana beberapa sub atribut untuk lebih mudah dipahami.

Dari hasil seleksi data tersebut dilakukan proses evaluasi menggunakan algoritma C4.5 dan diimplementasikan menggunakan tools Weka.

**b. Pengujian Algoritma C 4.5 Dengan Software WEKA**

**1. Hasil Klasifikasi algoritma C4.5 menggunakan tools Weka ( 2 Fold Cross Validation)**

```
Classifier output
Time taken to build model: 0.01 seconds
=== Stratified cross-validation ===
=== Summary ===
Correctly Classified Instances      100          70.2609 %
Incorrectly Classified Instances    30           21.7391 %
Kappa statistic                    0.6290
Mean absolute error                 0.1070
Root mean squared error             0.3424
Relative absolute error             45.0437 %
Root relative squared error         76.2343 %
Total Number of Instances          138

=== Detailed Accuracy By Class ===
          TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
0.946  0.158  0.854  0.946  0.897  0.770  0.955  0.847  Auditori
0.456  0.575  0.724  0.456  0.498  0.402  0.473  0.424  Visual
0.531  0.594  0.430  0.531  0.576  0.445  0.713  0.537  Kinestetik
Weighted Avg.  0.783  0.140  0.772  0.783  0.775  0.660  0.842  0.723

=== Confusion Matrix ===
 a b c <-- classified as
 70 3 1 | a = Auditori
 2 21 9 | b = Visual
 10 5 17 | c = Kinestetik
```

**Gambar 2.** Klasifikasi Algoritma C4.5 (2 Fold Cross Validation)

Sumber : Hasil Perhitungan

**Tabel 1.** Persentase Klasifikasi 2 Fold Cross Validation

Akurasi	Presisi	Recall
78%	77%	78%

Sumber : Hasil Perhitungan

**Tabel 2.** Confusion Matrix pada 2 Fold Cross Validation

		Predicted		
		Auditori	Visual	Kinestetik
Aktual	Auditori	70	3	1
	Visual	2	21	9
	Kinestetik	10	5	17

Sumber : Hasil Perhitungan

**2. Hasil Klasifikasi algoritma C4.5 menggunakan tools Weka (3 Fold Cross Validation)**

```
Classifier output
Time taken to build model: 0 seconds
=== Stratified cross-validation ===
=== Summary ===
Correctly Classified Instances      112          81.1594 %
Incorrectly Classified Instances    26           18.8406 %
Kappa statistic                    0.6790
Mean absolute error                 0.1853
Root mean squared error             0.3274
Relative absolute error             45.7797 %
Root relative squared error         73.8774 %
Total Number of Instances          138

=== Detailed Accuracy By Class ===
          TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
0.946  0.180  0.854  0.946  0.897  0.770  0.868  0.823  Auditori
0.544  0.584  0.730  0.544  0.703  0.714  0.586  0.723  Visual
0.469  0.508  0.759  0.469  0.588  0.528  0.731  0.574  Kinestetik
Weighted Avg.  0.812  0.131  0.810  0.812  0.799  0.701  0.841  0.742

=== Confusion Matrix ===
 a b c <-- classified as
 70 3 1 | a = Auditori
 2 27 3 | b = Visual
 10 7 15 | c = Kinestetik
```

**Gambar 3.** Klasifikasi Algoritma C4.5 (3 Fold Cross Validation)

Sumber : Hasil Perhitungan

**Tabel 3.** Persentase Klasifikasi 3 Fold Cross Validation

Akurasi	Presisi	Recall
81%	81%	81%

Sumber : Hasil Perhitungan

**Tabel 4.** Confusion Matrix pada 3 Fold Cross Validation

		Predicted		
		Auditori	Visual	Kinestetik
Aktual	Auditori	70	3	1
	Visual	2	27	3
	Kinestetik	10	7	15

Sumber : Hasil Perhitungan

**3. Hasil Klasifikasi algoritma C4.5 menggunakan tools Weka (4 Fold Cross Validation)**

```
Classifier output
Time taken to build model: 0 seconds
=== Stratified cross-validation ===
=== Summary ===
Correctly Classified Instances      111          80.4348 %
Incorrectly Classified Instances    27           19.5652 %
Kappa statistic                    0.6668
Mean absolute error                 0.201
Root mean squared error             0.3293
Relative absolute error             49.0921 %
Root relative squared error         73.3313 %
Total Number of Instances          138

=== Detailed Accuracy By Class ===
          TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
0.946  0.180  0.854  0.946  0.897  0.770  0.870  0.844  Auditori
0.675  0.113  0.700  0.675  0.778  0.709  0.892  0.642  Visual
0.406  0.528  0.813  0.406  0.542  0.498  0.737  0.571  Kinestetik
Weighted Avg.  0.804  0.133  0.808  0.804  0.787  0.693  0.848  0.734

=== Confusion Matrix ===
 a b c <-- classified as
 70 3 1 | a = Auditori
 2 28 2 | b = Visual
 10 9 13 | c = Kinestetik
```

**Gambar 4.** Klasifikasi Algoritma C4.5 (4 Fold Cross Validation)

Sumber : Hasil Perhitungan

**Tabel 5.** Persentase Klasifikasi 4 Cross Validation

Akurasi	Presisi	Recall
80%	80%	80%

Sumber : Hasil Perhitungan

**Tabel 6.** Confusion Matrix pada 4 Fold Cross Validation

		Predicted		
		Auditori	Visual	Kinestetik
Aktual	Auditori	70	3	1
	Visual	2	28	2
	Kinestetik	10	9	13

Sumber : Hasil Perhitungan

4. Hasil Klasifikasi algoritma C4.5 menggunakan tools Weka (5 Fold Cross Validation)

```
Classifier output
Time taken to build model: 0.03 seconds

--- Stratified cross-validation ---
--- Summary ---
Correctly Classified Instances      111      80.4348 %
Incorrectly Classified Instances    27       19.5652 %
Kappa statistic                    0.6668
Mean absolute error                 0.1996
Root mean squared error             0.3234
Relative absolute error             49.3455 %
Root relative squared error         74.2095 %
Total Number of Instances          138

--- Detailed Accuracy By Class ---
TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
0.946   0.188   0.854     0.946   0.897     0.770   0.856   0.817   Auditori
0.875   0.113   0.700     0.875   0.770     0.709   0.874   0.654   Visual
0.406   0.023   0.813     0.406   0.542     0.498   0.706   0.519   Kinestetik
Weighted Avg.  0.804   0.133   0.808     0.804   0.787     0.693   0.826   0.710

--- Confusion Matrix ---
a b c <-- classified as
70 3 1 | a = Auditori
2 28 2 | b = Visual
10 9 13 | c = Kinestetik
```

Gambar 5. Klasifikasi Algoritma C4.5 (5 Fold Cross Validation)

Sumber : Hasil Perhitungan

Tabel 7. Persentase Klasifikasi 5 Fold Cross Validation

Akurasi	Presisi	Recall
80%	80%	80%

Sumber : Hasil Perhitungan

Tabel 8. Confusion Matrix pada 5 Fold Cross Validation

		Predicted		
		Auditori	Visual	Kinestetik
Aktual	Auditori	70	3	1
	Visual	2	28	2
	Kinestetik	10	9	13

Sumber : Hasil Perhitungan

5. Hasil Klasifikasi Algoritma C4.5 Menggunakan Tools Weka (6 Fold Cross Validation)

```
Classifier output
Time taken to build model: 0.03 seconds

--- Stratified cross-validation ---
--- Summary ---
Correctly Classified Instances      111      80.4348 %
Incorrectly Classified Instances    27       19.5652 %
Kappa statistic                    0.6668
Mean absolute error                 0.1996
Root mean squared error             0.3234
Relative absolute error             49.3455 %
Root relative squared error         74.2095 %
Total Number of Instances          138

--- Detailed Accuracy By Class ---
TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
0.946   0.188   0.854     0.946   0.897     0.770   0.856   0.817   Auditori
0.875   0.113   0.700     0.875   0.770     0.709   0.874   0.654   Visual
0.406   0.023   0.813     0.406   0.542     0.498   0.706   0.519   Kinestetik
Weighted Avg.  0.804   0.133   0.808     0.804   0.787     0.693   0.826   0.710

--- Confusion Matrix ---
a b c <-- classified as
70 3 1 | a = Auditori
2 28 2 | b = Visual
10 9 13 | c = Kinestetik
```

Gambar 6. Klasifikasi Algoritma C4.5 (6 Fold Cross Validation)

Sumber : Hasil Perhitungan

Tabel 9. Persentase Klasifikasi 6 Fold Cross Validation

Akurasi	Presisi	Recall
79%	78%	79%

Sumber : Hasil Perhitungan

Tabel 10. Confusion Matrix pada 6 Fold Cross Validation

		Predicted		
		Auditori	Visual	Kinestetik
Aktual	Auditori	70	3	1
	Visual	2	26	4
	Kinestetik	10	9	13

Sumber : Hasil Perhitungan

6. Hasil Klasifikasi Algoritma C4.5 Menggunakan Tools Weka (7 Fold Cross Validation)

```
Classifier output
Time taken to build model: 0 seconds

--- Stratified cross-validation ---
--- Summary ---
Correctly Classified Instances      111      80.4348 %
Incorrectly Classified Instances    27       19.5652 %
Kappa statistic                    0.6668
Mean absolute error                 0.2001
Root mean squared error             0.3236
Relative absolute error             49.4634 %
Root relative squared error         73.347 %
Total Number of Instances          138

--- Detailed Accuracy By Class ---
TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
0.946   0.188   0.854     0.946   0.897     0.770   0.863   0.830   Auditori
0.906   0.123   0.690     0.906   0.784   0.719   0.881   0.705   Visual
0.375   0.019   0.857     0.375   0.522     0.498   0.717   0.521   Kinestetik
Weighted Avg.  0.804   0.133   0.817     0.804   0.784   0.695   0.833   0.729

--- Confusion Matrix ---
a b c <-- classified as
70 3 1 | a = Auditori
2 29 1 | b = Visual
10 10 12 | c = Kinestetik
```

Gambar 7. Klasifikasi Algoritma C4.5 (7 Fold Cross Validation)

Sumber : Hasil Perhitungan

Tabel 11. Persentase Klasifikasi 7 Fold Cross Validation

Akurasi	Presisi	Recall
80%	81%	80%

Sumber : Hasil Perhitungan

Tabel 12. Confusion Matrix pada 7 Fold Cross Validation

		Predicted		
		Auditori	Visual	Kinestetik
Aktual	Auditori	70	3	1
	Visual	2	29	1
	Kinestetik	10	10	12

Sumber : Hasil Perhitungan

7. Hasil Klasifikasi Algoritma C4.5 Menggunakan *Tools Weka* (8 Fold Cross Validation)

```
Classifier output
Time taken to build model: 0 seconds
--- Stratified cross-validation ---
--- Summary ---
Correctly Classified Instances      111      80.4348 %
Incorrectly Classified Instances    27       19.5652 %
Kappa statistic                    0.6668
Mean absolute error                0.202
Root mean squared error            0.3382
Relative absolute error            49.7233 %
Root relative squared error        73.0941 %
Total Number of Instances         139

--- Detailed Accuracy By Class ---
TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
0.946  0.188  0.854  0.946  0.897  0.770  0.864  0.830  Auditori
0.875  0.113  0.700  0.875  0.778  0.709  0.870  0.673  Visual
0.406  0.528  0.513  0.406  0.542  0.498  0.709  0.517  Kinestetik
Weighted Avg.  0.804  0.133  0.508  0.804  0.787  0.693  0.829  0.721

--- Confusion Matrix ---
a b c <-- classified as
70 3 1 | a = Auditori
2 28 2 | b = Visual
10 9 13 | c = Kinestetik
```

**Gambar 8.** Klasifikasi Algoritma C4.5 (8 Cross Validation)

Sumber : Hasil Perhitungan  
**Tabel 13.** Persentase Klasifikasi 8 Fold Cross Validation

Akurasi	Presisi	Recall
80%	80%	80%

Sumber : Hasil Perhitungan  
**Tabel 14.** Confusion Matrix pada 8 Fold Cross Validation

		Predicted		
		Auditori	Visual	Kinestetik
Aktual	Auditori	70	3	1
	Visual	2	28	2
	Kinestetik	10	9	13

Sumber : Hasil Perhitungan  
 8. Hasil Klasifikasi Algoritma C4.5 Menggunakan *Tools Weka* (9 Fold Cross Validation)

```
Classifier output
Time taken to build model: 0 seconds
--- Stratified cross-validation ---
--- Summary ---
Correctly Classified Instances      110      79.1101 %
Incorrectly Classified Instances    29      20.8899 %
Kappa statistic                    0.6545
Mean absolute error                0.1996
Root mean squared error            0.3354
Relative absolute error            49.3543 %
Root relative squared error        74.634 %
Total Number of Instances         139

--- Detailed Accuracy By Class ---
TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
0.946  0.188  0.854  0.946  0.897  0.770  0.853  0.810  Auditori
0.875  0.113  0.693  0.875  0.767  0.698  0.851  0.667  Visual
0.375  0.508  0.500  0.375  0.511  0.470  0.655  0.501  Kinestetik
Weighted Avg.  0.797  0.136  0.502  0.797  0.778  0.683  0.813  0.705

--- Confusion Matrix ---
a b c <-- classified as
70 3 1 | a = Auditori
2 28 2 | b = Visual
10 10 12 | c = Kinestetik
```

**Gambar 9.** Klasifikasi Algoritma C4.5 (9 Cross Validation)

Sumber : Hasil Perhitungan  
**Tabel 15.** Persentase Klasifikasi 9 Fold Cross Validation

Akurasi	Presisi	Recall
80%	80%	80%

Sumber : Hasil Perhitungan  
**Tabel 16.** Confusion Matrix pada 9 Fold Cross Validation

		Predicted		
		Auditori	Visual	Kinestetik
Aktual	Auditori	70	3	1
	Visual	2	28	2
	Kinestetik	10	10	12

Sumber : Hasil Perhitungan  
 9. Hasil Klasifikasi Algoritma C4.5 Menggunakan *Tools Weka* (10 Fold Cross Validation)

```
Classifier output
Time taken to build model: 0.01 seconds
--- Stratified cross-validation ---
--- Summary ---
Correctly Classified Instances      111      80.4348 %
Incorrectly Classified Instances    27       19.5652 %
Kappa statistic                    0.6668
Mean absolute error                0.2028
Root mean squared error            0.3382
Relative absolute error            50.1317 %
Root relative squared error        74.6171 %
Total Number of Instances         139

--- Detailed Accuracy By Class ---
TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
0.946  0.188  0.854  0.946  0.897  0.770  0.864  0.816  Auditori
0.875  0.113  0.700  0.875  0.778  0.709  0.870  0.624  Visual
0.406  0.528  0.513  0.406  0.542  0.498  0.693  0.490  Kinestetik
Weighted Avg.  0.804  0.133  0.508  0.804  0.787  0.693  0.815  0.696

--- Confusion Matrix ---
a b c <-- classified as
70 3 1 | a = Auditori
2 28 2 | b = Visual
10 9 13 | c = Kinestetik
```

**Gambar 10.** Klasifikasi Algoritma C4.5 (10 Cross Validation)

Sumber : Hasil Perhitungan  
**Tabel 17.** Persentase Klasifikasi 10 Fold Cross Validation

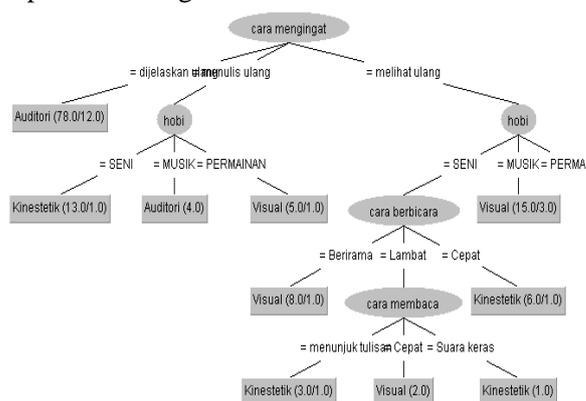
Akurasi	Presisi	Recall
80%	80%	80%

Sumber : Hasil Perhitungan  
**Tabel 18.** Confusion Matrix pada 10 Fold Cross Validation

		Predicted		
		Auditori	Visual	Kinestetik
Aktual	Auditori	70	3	1
	Visual	2	28	2
	Kinestetik	10	9	13

Sumber : Hasil Perhitungan

Berdasarkan hasil evaluasi menggunakan *cross validation* menghasilkan pohon keputusan sebagai berikut :



**Gambar 11.** Pohon Keputusan

Sumber : Hasil Perhitungan

Berikut adalah *rule* dari hasil pohon keputusan pada gambar 4.11

IF Cara Mengingat “Dijelaskan Ulang” then Gaya Belajar “Visual”

Else

IF Cara Mengingat “Menulis Ulang” and Hobi “Seni” then Gaya Belajar “Kinestetik

Else

IF Cara Mengingat "Menulis Ulang" and Hobi "Musik" then Gaya belajar "Auditori"

Else

IF Cara Mengingat "Menulis Ulang" and Hobi "Permainan" then Gaya belajar "Visual"

Else

IF Cara Mengingat "Melihat Ulang" and Hobi "Seni" and Cara Berbicara "Berirama" then Gaya Belajar "Visual"

Else

IF Cara Mengingat "Melihat Ulang" and Hobi "Seni" and Cara Berbicara "Lambat" and Cara Membaca "Menunjuk Tulisan" then Gaya Belajar "Kinestetik"

Else

IF Cara Mengingat "Melihat Ulang" and Hobi "Seni" and Cara Berbicara "Lambat" and Cara Membaca "Cepat" then Gaya Belajar "Visual"

Else

IF Cara Mengingat "Melihat Ulang" and Hobi "Seni" and Cara Berbicara "Lambat" and Cara Membaca "Suara Keras" then Gaya Belajar "Kinestetik"

Else

IF Cara Mengingat "Melihat Ulang" and Hobi "Seni" and Cara Berbicara "Cepat" then Gaya Belajar "Kinestetik"

Else

IF Cara Mengingat "Melihat Ulang" and Hobi "Musik" then Gaya Belajar "Visual"

Else

IF Cara Mengingat "Melihat Ulang" and Hobi "Permainan" then Gaya Belajar "Visual"

### c. Hasil Perbandingan Evaluasi Akurasi Algoritma C4.5 Dari Data Kuisisioner Siswa

Analisis klasifikasi algoritma C4.5 pada *tools Weka* menggunakan 2 *Fold Cross Validation*, 3 *Fold Cross Validation*, dan 4 *Cross Validation*, 5 *cross validation*, 6 *cross validation*, 7 *cross validation*, 8 *cross validation*, 9 *cross validation*, dan 10 *cross validation*, maka diperoleh akurasi tertinggi yaitu dengan 3 *Fold Cross Validation* dengan persentasi akurasi yaitu 81,1594% untuk prediksi benar dan 18,8406% untuk prediksi salah.

**Tabel 19.** Perbandingan Evaluasi Akurasi Algoritma C4.5

2 Fold Cross Validation	Correctly Classified Instances	78.28%
	Incorrectly Classified Instances	21.73%
3 Fold Cross Validation	Correctly Classified Instances	81.15%
	Incorrectly Classified Instances	18.84%
4 Fold Cross Validation	Correctly Classified Instances	80.43%
	Incorrectly Classified Instances	19.56%
5 Fold Cross Validation	Correctly Classified Instances	80.43%
	Incorrectly Classified Instances	19.57%
6 Fold Cross Validation	Correctly Classified Instances	78.99%
	Incorrectly Classified Instances	21.01%
7 Fold Cross Validation	Correctly Classified Instances	80.43%
	Incorrectly Classified Instances	19.57%
8 Fold Cross Validation	Correctly Classified Instances	80.43%
	Incorrectly Classified Instances	19.57%
9 Fold Cross Validation	Correctly Classified Instances	79.71%
	Incorrectly Classified Instances	20.29%
10 Fold Cross Validation	Correctly Classified Instances	80.43%
	Incorrectly Classified Instances	19.57%

Sumber : Hasil Perhitungan

### d. Faktor – Faktor Yang Mempengaruhi Gaya Belajar Berdasarkan Evaluasi Menggunakan Algoritma C4.5

Pada gaya belajar visual faktor yang mempengaruhi yang pertama adalah cara mengingat dengan menulis ulang dan hobi permainan, untuk faktor kedua dari cara mengingat melihat ulang hobi seni acara berbicara berirama, faktor ketiga yaitu cara membaca dengan cepat, dan yang terakhir cara mengingat melihat ulang hobi musik dan permainan.

Faktor yang mempengaruhi gaya belajar kinestetik adalah cara mengingat menulis ulang dan hobi seni, faktor kedua yaitu cara mengingat

melihat ulang, hobi seni cara berbicara lambat cara membaca menunjuk tulisan, dan suara keras, faktor yang terakhir yaitu dengan cara membaca yang cepat. Faktor yang mempengaruhi gaya belajar auditori adalah dengan cara mengingat dijelaskan ulang, cara mengingat dengan menulis ulang dan hobi music.

## 5. KESIMPULAN DAN SARAN

### a. Kesimpulan

Berdasarkan hasil pengujian klasifikasi menggunakan algoritma C4.5 pada dataset kuisisioner siswa-siswi di SMP Negeri 2 Mayang dengan jumlah data sebanyak 138 data dan menggunakan 6 atribut untuk membantu proses klasifikasi yaitu metode belajar, hobi, cara membaca, cara mengingat, cara berbicara, cara memahami dan dengan kelas atribut gaya belajar. Perhitungan menggunakan metode *cross validation* menghasilkan rata-rata nilai *Correctly Classified Instances* 80,03% dan *Incorrectly Classified Instances* 19,97%. Sedangkan faktor - faktor yang mempengaruhi gaya belajar siswa sesuai hasil klasifikasi dilihat dari cara mengingat, cara membaca, cara berbicara dan hobi.

### b. Saran

Bagi peneliti-peneliti selanjutnya yang akan melakukan penelitian serupa dan mengembangkan penelitian ini yaitu:

1. Data perlu menyesuaikan dengan kurikulum yang terbaru.
2. Sebaiknya memperbanyak jumlah data kuisisioner ditambah, sehingga dapat memperoleh hasil akurasi yang lebih baik
3. Diperlukan tingkat ketelitian yang tinggi dalam pemilihan data agar tidak terjadi *Error* dalam pemrosesan data dalam *tools Weka*.

## 6. DAFTAR PUSTAKA

- Atmaja, K. J., Anandita, I. B. G., & Dewi, N. K. C. (2018). Penerapan Data Mining Untuk Memprediksi Potensi Pendonor Darah Menjadi Pendonor Tetap Menggunakan Metode Decision Tree C.45. *S@Cies*, 7(2), 101–108. <https://doi.org/10.31598/sacies.v7i2.284>
- Azwanti, N. (2018). Analisa Algoritma C4.5 Untuk Memprediksi Penjualan Motor Pada Pt. Capella Dinamik Nusantara Cabang Muka Kuning. *Informatika Mulawarman: Jurnal Ilmiah Ilmu Komputer*, 13(1), 33. <https://doi.org/10.30872/jim.v13i1.629>
- Elisa, E. (2017). Analisa dan Penerapan Algoritma C4.5 Dalam Data Mining Untuk Mengidentifikasi Faktor-Faktor Penyebab Kecelakaan Kerja Kontruksi PT.Arupadhatu Adisesanti. *Jurnal Online Informatika*, 2(1), 36. <https://doi.org/10.15575/join.v2i1.71>
- GINANJAR, A. R. (2019). *Sistem Deteksi Jenis Cacat Biji Kopi dengan Algoritma K-Nearest Neighbor PROGRAM STUDI TEKNIK INFORMATIKA*.
- Informasi, J. S., & Informatika, J. T. (2005). *definisi datamining*. 2–4.
- Kadir, F., Permana, I., & Qalby, N. (2020). Pengaruh Gaya Belajar Siswa Terhadap Hasil Belajar Fisika Sma Pgri Maros. *Karst: JURNAL PENDIDIKAN FISIKA DAN TERAPANNYA*, 3(1), 91–95. <https://doi.org/10.46918/karst.v3i1.538>
- Lukhayu Pritalia, G. (2018). Penerapan Algoritma C4.5 untuk Penentuan Ketersediaan Barang E-commerce. *Indonesian Journal of Information Systems*, 1(1), 47–56. <https://doi.org/10.24002/ijis.v1i1.1727>
- Novianti, T., & Santosa, I. (2016). PENENTUAN JADWAL KERJA BERDASARKAN KLASIFIKASI DATA KARYAWAN MENGGUNAKAN METODE DECISION TREE C4.5 (Studi Kasus Universitas Muhammadiyah Surabaya). *Jurnal Komunika: Jurnal Komunikasi, Media Dan Informatika*, 5(1), 1. <https://doi.org/10.31504/komunika.v5i1.633>
- Nugraha, P. G. S. C., Aribawa, I. W., Priyana, I. P. O., & Indrawan, G. (2016). Penerapan Metode Decision Tree(Data Mining) Untuk Memprediksi Tingkat Kelulusan Siswa Smpn1 Kintamani. *Seminar Nasional Vokasi Dan Teknologi (SEMNASVOKTEK)*, 35–44.
- Prayoga, N. D. (2018). *Penerapan Algoritma C.45 Dalam Memprediksi Kelulusan*

- Tepat Waktu Pada Perguruan Tinggi  
(Studi Kasus : Smik Royal Kisaran).*  
<https://doi.org/10.31227/osf.io/unqt4>
- Rani, L. N. (2016). Klasifikasi Nasabah Menggunakan Algoritma C4.5 Sebagai Dasar Pemberian Kredit. *INOVTEK Polbeng - Seri Informatika*, 1(2), 126.  
<https://doi.org/10.35314/isi.v1i2.131>