

# KLASIFIKASI SISWA BERDASARKAN MATA PELAJARAN LINTAS MINAT MENGUNAKAN METODE DECISION TREE C4.5

Suherman<sup>1</sup>, Marlia Purnamasari<sup>2</sup>, Fitriani Dwi Hastuti<sup>3</sup>

<sup>1,2,3</sup> *Jurusan Teknik Informatika Fakultas Teknologi Informasi Universitas Serang Raya  
Jln. Raya Cilegon Serang – Drangong Kota Serang*

<sup>1</sup>suherman.unsera@gmail.com

<sup>2</sup>marlia.unsera@gmail.com

<sup>3</sup>fitrianiidwihastuti@gmail.com

**Abstrak** - Kurikulum 2013 dirancang untuk memberikan kesempatan kepada siswa belajar berdasarkan minat siswa. Selain memilih mata pelajaran dalam suatu peminatan tertentu, siswa juga diberi kesempatan untuk mengambil mata pelajaran lintas minat. SMA Negeri 1 Anyer salah satu sekolah yang telah menerapkan program lintas minat. Dalam proses penentuan kelas lintas minat disekolah tersebut masih mengalami kendala yaitu tidak terspesifikasinya siswa yang memiliki minat pada mata pelajaran tertentu, dan pada proses pemilihan lintas minat ditentukan oleh pihak sekolah. Penelitian ini bertujuan untuk mengklasifikasi siswa berdasarkan minat dan bakat siswa pada mata pelajaran tertentu. Metode yang digunakan yaitu *Decision Tree* dan algoritma C4.5. Pada penelitian ini didapat nilai akurasi sebesar 82,82%. Penelitian menghasilkan sebuah sistem penentuan kelas lintas Minat. Model klasifikasi ini dapat membantu siswa dalam menentukan lintas minat dan dapat digunakan sebagai alternatif referensi bagi guru BK untuk dapat mengelompokkan siswa berdasarkan minat dan bakat siswa.

**Kata Kunci** : Algoritma C4.5, *Decision Tree*, Klasifikasi, Lintas Minat

## I. PENDAHULUAN

Kurikulum Tahun 2013 dirancang untuk memberikan kesempatan kepada siswa belajar berdasarkan minat mereka. Struktur kurikulum tahun 2013 menyediakan mata pelajaran wajib yang diikuti oleh seluruh peserta didik di satu satuan pendidikan pada setiap satuan dan jenjang pendidikan dan mata pelajaran pilihan yang diikuti oleh peserta didik sesuai dengan pilihan mereka.

Dalam Kurikulum 2013, selain memilih mata pelajaran dalam suatu peminatan tertentu, siswa juga diberi kesempatan untuk mengambil mata pelajaran antar kelompok peminatan atau yang disebut lintas minat. Lintas Minat adalah program kulikuler yang disediakan untuk mengakomodasi perluasan pilihan minat, bakat dan/atau kemampuan akademik siswa dengan orientasi penguasaan kelompok mata pelajaran keilmuan di luar pilihan minat. Hal ini memberi peluang kepada siswa untuk mempelajari mata pelajaran yang diminati namun tidak terdapat pada kelompok mata pelajaran peminatan.

Lintas minat adalah program untuk memperluas dan mengembangkan minat, bakat dan kemampuan peserta didik yang mereka miliki dengan memilih kelompok mata pelajaran, di luar kelompok program peminatannya [1]. Menurut Innike Marbitha Putri [2] Program ini bertujuan untuk memberikan kesempatan dan peluang kepada siswa untuk dapat memilih dan mempelajari mata pelajaran yang tidak ada pada program peminatan.

SMA Negeri 1 Anyer merupakan salah satu lembaga pendidikan menengah negeri di kabupaten Serang. Sesuai dengan keputusan pemerintah No. 81A Tahun 2013 tentang Implementasi Kurikulum 2013, maka SMA Negeri 1 Anyer sesuai dengan keputusan tersebut juga wajib menerapkan kurikulum 2013. Dalam proses penentuan mata pelajaran lintas minat pada sekolah tersebut masih mengalami kendala yaitu tidak terspesifikasinya siswa yang memiliki minat pada mata pelajaran tertentu dan pada proses pemilihan kelas lintas minat ditentukan oleh pihak sekolah. Jika kelas lintas minat tidak sesuai dengan minat dan kemampuan siswa, maka akan sangat berpengaruh pada proses belajar siswa kedepannya.

Penelitian ini akan memberikan kemudahan bagi siswa untuk mengetahui kelas lintas minat yang sesuai dengan minat dan bakat, yaitu dengan pengembangan aplikasi menggunakan metode *decision tree* dan Algoritma C4.5, dengan memberikan pilihan – pilihan kriteria mata pelajaran lintas minat maka siswa akan memilih sesuai dengan minat dan bakat siswa tersebut.

Gorunescu mengemukakan bahwa *decision tree* atau pohon keputusan adalah pohon yang digunakan sebagai prosedur penalaran untuk mendapatkan jawaban dari masalah yang dimasukkan. Kefleksibelan membuat metode ini atraktif, khususnya karena memberikan keuntungan berupa visualisasi saran (dalam bentuk *decision tree*) yang membuat prosedur prediksinya dapat diamati [3].

Penelitian ini akan membahas model data mining yang dapat memberikan solusi penyelesaian atas kendala siswa yang akan memilih mata pelajaran lintas minat dan pengembangan aplikasi ini berbasis web. Cara kerja aplikasi ini yaitu user memilih salah satu dari kriteria yang telah disediakan. Hasil penentuan mata pelajaran lintas minat didasarkan atas hasil perhitungan dengan metode klasifikasi *decision tree* dan menggunakan algoritma C4.5. Jika hasil dari perhitungan tersebut cenderung pada mata pelajaran pertama, maka siswa tersebut disarankan untuk masuk ke dalam kelas mata pelajaran pertama. Begitupun jika hasil perhitungannya cenderung pada mata pelajaran kedua atau ketiga maka siswa tersebut disarankan untuk masuk dalam kelas mata pelajaran kedua atau ketiga.

Pada penelitian ini menghasilkan sebuah aplikasi yang dikembangkan dengan bahasa pemrograman PHP sebagai *Server Side Programming* dan MySQL sebagai *Database Server*. Dalam perancangan sistemnya menggunakan UML (*Unified Modeling Language*), yaitu sekumpulan konversi pemodelan yang digunakan untuk menentukan atau menggambarkan sebuah sistem software yang terkait dengan objek [4]. UML berarti memiliki sintaks dan sistematis. Ketika membuat model menggunakan konsep UML ada aturan-aturan yang harus diikuti. Bagaimana elemen pada model-model yang dibuat berhubungan satu dengan lainnya harus mengikuti standar yang ada. Seperti *use case diagram*, *activity diagram*, *sequence diagram* dan *diagram class* [5].

## II. METODOLOGI PENELITIAN

Sumber data adalah segala sesuatu yang dapat memberikan informasi mengenai data. Berdasarkan sumbernya, data dibedakan menjadi dua yaitu data primer dan sekunder [6]. Pada penelitian ini sumber data yang digunakan oleh peneliti yaitu sumber data primer yang didapat melalui kuesioner pada siswa langsung.

Data mining adalah suatu istilah yang digunakan untuk menguraikan penemuan pengetahuan di dalam *database*. Data mining adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan machine learning untuk mengekstraksi dan mengidentifikasi informasi yang

bermanfaat dan pengetahuan yang terkait dari berbagai database besar [7].

### 2.1 Decision Tree

*Decision Tree* atau pohon keputusan merupakan grafik diagram alir yang mewakili proses pengambilan keputusan, dimana grafik diagram alir tersebut menyerupai bentuk pohon. Pohon keputusan dapat digunakan seseorang untuk menentukan keputusan yang sulit dengan menyederhanakannya ke dalam pilihan yang lebih mudah. Setiap pohon keputusan memiliki simpul (*node*) dan cabang (*branch*) yang menghubungkan setiap simpul (*nodes*). Bagian simpul yang terletak di bagian bawah pohon keputusan disebut dengan kelas keputusan (*leaves*) sedangkan simpul paling atas dari pohon keputusan disebut dengan root. Melalui root inilah dapat diketahui keseluruhan sample data training yang sudah terbagi ke dalam klas-klas keputusan tertentu. Setiap simpul pada pohon keputusan (kecuali *leaves*) disebut sebagai simpul keputusan. Simpul keputusan inilah yang memberikan informasi keputusan berdasarkan fitur tunggal berupa value yang merujuk pada salah satu *leaves* yang dimilikinya [8].

### 2.2 Algoritma C4.5

C4.5 diperkenalkan oleh Quinlan (1996) sebagai versi perbaikan dari ID3. Perbaikan yang membedakan algoritma C4.5 dari ID3 adalah dapat menangani fitur dengan tipe numerik, melakukan pemotongan (*pruning*) *decision tree*, dan penurunan (*deriving*) *rule set*. Algoritma C4.5 juga menggunakan kriteria gain dalam menentukan fitur yang menjadi pemecah *node* pada pohon yang diinduksi [3].

Berikut langkah-langkah penyelesaian *decision tree* menggunakan algoritma C4.5 [8]:

#### 1. *Developing and understanding the application domain*

Tujuan yang ingin dicapai dari proses ini yaitu klasifikasi siswa berdasarkan lintas minat.

#### 2. *Creating a target data set*

Variabel (atribut) yang digunakan yaitu nama, nispn, jenis\_kelamin, kelas, jawaban\_a, jawaban\_b, dan jawaban\_c. Data *point (sample)* yang akan digunakan yaitu minat\_asli.

#### 3. *Data cleaning and preprocessing*

Pada tahap ini dilakukan proses penghapusan data yang tidak diharapkan keberadaannya (*outliers*), atribut yang dihilangkan yaitu: nispn, jenis\_kelamin dan kelas.

#### 4. *Data reduction and projection*

Dari tahapan data cleaning and preprocessing maka atribut yang digunakan yaitu jawaban\_a, jawaban\_b dan jawaban\_c. Ketiga atribut tersebut merupakan kriteria numerik/kontinyu. Untuk meningkatkan tingkat akurasi metode klasifikasi, maka dilakukan proses diskritisasi menggunakan teknik *unsupervised discretization*.

Diantara metode *unsupervised Discretization*, terdapat beberapa metode yang sederhana. Seperti (*equal-width Interval Discretization* dan *equal- frequency Interval*

Discretization) dan yang lebih canggih, berdasarkan analisis clustering, seperti k-means discretization. Rentang kontinu dibagi menjadi sub rentang oleh lebar yang ditentukan pengguna atau Frekuensi [9].

Teknik unsupervised discretization pada kriteria yang bernilai numerik/kontinu, tujuannya adalah untuk mentransformasikan kriteria yang bernilai numerik/kontinu menjadi kriteria kategorikal seperti berikut :

$$Interval = \frac{V_{max} - V_{min}}{k}$$

$$Interval = \frac{30 - 0}{3} = 10$$

Kemudian menentukan batas-batas sebagai berikut :

$$Boundaries = V_{min} + (i * interval)$$

$$Boundaries (1) = 0 + (1 * 10) = 10$$

$$Boundaries (2) = 0 + (2 * 10) = 20$$

$$Boundaries (3) = 0 + (3 * 10) = 30$$

Maka diperoleh nilai kategorikal atribut sebagai berikut :

TABEL I. NILAI KATEGORIKAL

Jawaban_A	Jawaban_B	Jawaban_C
≤10	≤10	≤10
>10	>10	>10
≤20	≤20	≤20
>20	>20	>20
≤30	≤30	≤30
>30	>30	>30

5. Choosing the data mining task

Metode data mining yang digunakan yaitu pohon klasifikasi, karena sesuai dengan tujuan yang ingin dicapai.

6. Choosing the data mining algorithm

Algoritma yang digunakan yaitu algoritma C4.5.

7. Data mining

TABEL II. DATA TRAINING PERTAMA

No	Nama	A	B	C	Minat Asli
1	Abdul Malik R	18	9	3	Biologi
2	Achmad Daviq	8	10	12	Fisika
3	Adnan Setiawan	11	11	8	Biologi
4	Adrian Maulana	13	9	8	Biologi
5	Agus Irawan	5	10	15	Fisika
6	Ahmad As'adi Z	5	13	12	Kimia
7	Ahmad Fauzi S	9	9	12	Fisika
8	Ahmad Nugri	12	10	8	Biologi
9	Aldi Yusufandi	11	4	15	Fisika
10	Alif Riswanda	14	7	9	Biologi
11	Alifia Najwa A	14	9	7	Biologi
12	Alvia Septiani	17	12	1	Biologi
13	Alvin Maulana	19	6	5	Biologi
14	Ananda Widya	11	9	10	Biologi
15	Anes Awaliyah	12	15	3	Kimia
16	Anisa S	15	9	6	Biologi
17	Arif F	9	14	7	Kimia
18	Arya Darma J.S	13	12	5	Biologi
19	Arya Difa	12	10	8	Biologi

20	Arya Dwi S	17	6	7	Biologi
21	Asnah	25	3	2	Biologi
22	Bagas Parmono	12	11	7	Biologi
23	Cindy Ayu K S	8	19	3	Kimia
24	Cinta Azahra	25	2	3	Biologi
25	Danu S	21	6	3	Biologi
26	Desti Vinita A	12	7	11	Biologi
27	Dewi Ratnasari	14	10	6	Biologi
28	Dhayanu P	23	2	5	Biologi
29	Dhea Liana P	17	10	3	Biologi
30	Diki Rosadi	5	12	13	Fisika
31	Dimas P. A	10	10	10	Fisika
32	Dini Wahyuni	17	10	3	Biologi
33	Dita Imelda	13	12	5	Biologi
34	Endar Rifaldi	12	11	7	Biologi
35	Eri Irawan	14	12	4	Biologi
36	Erlinda Putri	14	7	9	Biologi
37	Fadila R	7	19	4	Kimia
38	Faizal Bagus Z	9	12	9	Kimia
39	Farid Puji F R	6	9	15	Fisika
40	Fariz Ilham R	9	12	9	Kimia
41	Fasya Saskia	11	12	7	Kimia
42	Fauziah	22	6	2	Biologi
43	Fironica N. A	16	9	5	Biologi
44	Hafidz Ainur R	14	8	8	Biologi
45	Hani Hunarti	11	13	6	Kimia
46	Ihya Ulum M	7	6	17	Fisika
47	Ike Mutiara A	19	8	3	Biologi
48	Indra Maulana	14	6	10	Biologi
49	Intan Wulandari	15	7	8	Biologi
50	Jahratul H	26	3	1	Biologi
51	Jimi Fauzan	8	9	12	Fisika
52	Juhiar	9	12	9	Kimia
53	Julita	18	7	4	Biologi
54	Kayla Rayane P	14	11	5	Biologi
55	Kholis Femas F	13	14	3	Kimia
56	Lis Nurmala	12	13	5	Kimia
57	Lisna N	12	15	3	Kimia
58	M Fikri Rivaldi	8	10	12	Fisika
59	M Nascar R	16	8	6	Biologi
60	M Rojid	7	17	8	Kimia
61	Marcella Devita	12	13	5	Kimia
62	Mario Maulana	30	0	0	Biologi
63	Marni	12	13	5	Kimia
64	Masani	4	23	3	Kimia
65	Maudy Al	17	10	3	Biologi
66	Maulinda P N	16	6	8	Biologi
67	Mohamad Dafa	10	14	6	Kimia
68	Muhammad C S	11	6	13	Fisika
69	Muhammad Z	12	10	8	Biologi
70	Mulyana Sari	30	0	0	Biologi
71	Nabila Novianti	16	9	5	Biologi
72	Nafis Baldan	16	8	6	Biologi
73	Nasywa Dany A	12	10	8	Biologi
74	Nessa Eka Putri	13	7	10	Biologi
75	Nilam Cahya	16	10	4	Biologi
76	Nisa Zahrani	2	3	25	Fisika
77	Nur Ayu P	12	13	5	Kimia
78	Nurchayani	11	7	12	Fisika
79	Nurjanah	8	20	2	Kimia
80	Nurul Sofiana	24	5	1	Biologi
81	Raditya Dwi P	10	7	13	Fisika
82	Raisha Azzahra	15	8	7	Biologi
83	Rendi Setiawan	18	5	7	Biologi
84	Rikaz Ayatullah	12	11	7	Biologi

85	Riska Agustina	10	12	8	Kimia
86	Riska Utami	14	10	6	Biologi
87	Rizka Alifia M	16	8	6	Biologi
88	Salwa Saidah A	22	5	3	Biologi
89	Serlina	4	15	11	Kimia
90	Shabilla Siti R	24	6	0	Biologi
91	Siti Nurdiana A	14	13	3	Biologi
92	Siti Nurma A	6	7	17	Fisika
93	Siti Salma S	19	9	2	Biologi
94	Siti Sya'riatul M	5	10	15	Fisika
95	Sri Mulliyati	20	9	1	Biologi
96	Syifa Aulia R G	12	11	7	Biologi
97	Trisda Cahya R	20	6	4	Biologi
98	Valentina	9	9	12	Fisika
99	Wilda Meilani	19	8	3	Biologi
100	Windi W	16	12	2	Biologi
101	Yoshavath A P	11	11	8	Kimia
102	Yunita Fitria D	27	2	1	Biologi
103	Zahra Rizki A	11	11	8	Biologi

Sumber : SMA Negeri 1 Anyer, 2019

Penyelesaian

a. Tentukan jumlah himpunan Biologi, Kimia dan Fisika  
Pada kasus ini jumlah himpunan biologi yakni sebanyak 63, kimia sebanyak 23 dan fisika sebanyak 17.

b. Hitung nilai *entropy sample training* total

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

$$Entropy(total) = -\frac{63}{103} \log_2 \left(\frac{63}{103}\right) - \frac{23}{103} \log_2 \left(\frac{23}{103}\right) - \frac{17}{103} \log_2 \left(\frac{17}{103}\right) = 1,346$$

c. Hitung nilai *entropy* setiap atribut terhadap *entropy* total.  
Perhitungan *Entropy* dilakukan terhadap kelas atribut *Decision*.

1) Menghitung nilai *entropy* jawaban\_a ≤10 dan >10

Atribut jawaban\_a ≤10 dan >10 memiliki 2 buah jenis *value* yaitu ≤10 sebanyak 27 buah dan >10 sebanyak 76 buah. Dengan demikian perhitungan nilai *entropy* untuk setiap *value* atribut diperoleh sebagai berikut :

$$Entropy(S, jawaban_a \leq 10) = -\frac{13}{27} \log_2 \left(\frac{13}{27}\right) - \frac{14}{27} \log_2 \left(\frac{14}{27}\right) = 0,999$$

$$Entropy(S, jawaban_a > 10) = -\frac{63}{76} \log_2 \left(\frac{63}{76}\right) - \frac{10}{76} \log_2 \left(\frac{10}{76}\right) - \frac{3}{76} \log_2 \left(\frac{3}{76}\right) = 0,498$$

2) Menghitung nilai *entropy* jawaban\_a ≤20 dan >20

Atribut jawaban\_a ≤20 dan >20 memiliki 2 buah jenis *value* yaitu ≤20 sebanyak 91 buah dan >20 sebanyak 12. Dengan demikian perhitungan nilai *entropy* untuk setiap *value* atribut diperoleh sebagai berikut :

$$Entropy(S, jawaban_a \leq 20) = -\frac{51}{91} \log_2 \left(\frac{51}{91}\right) - \frac{23}{91} \log_2 \left(\frac{23}{91}\right) - \frac{17}{91} \log_2 \left(\frac{17}{91}\right) = 1,422$$

$$Entropy(S, jawaban_a > 20) = -\frac{12}{12} \log_2 \left(\frac{12}{12}\right) = 0$$

3) Menghitung nilai *entropy* jawaban\_a ≤30 dan >30

Atribut jawaban\_a ≤30 dan >30 memiliki 2 buah jenis *value* yaitu ≤30 sebanyak 103 buah dan >30 sebanyak 0

buah. Dengan demikian perhitungan nilai *entropy* untuk setiap *value* atribut diperoleh sebagai berikut :

$$Entropy(S, jawaban_a \leq 30) = -\frac{63}{103} \log_2 \left(\frac{63}{103}\right) - \frac{23}{103} \log_2 \left(\frac{23}{103}\right) - \frac{17}{103} \log_2 \left(\frac{17}{103}\right) = 1,346$$

$$Entropy(S, jawaban_a > 30) = -\frac{0}{0} \log_2 \left(\frac{0}{0}\right) = 0$$

4) Menghitung nilai *entropy* jawaban\_b ≤10 dan >10

Atribut jawaban\_b ≤10 dan >10 memiliki 2 buah jenis *value* yaitu ≤10 sebanyak 66 buah dan >10 sebanyak 37 buah. Dengan demikian perhitungan nilai *entropy* untuk setiap *value* atribut diperoleh sebagai berikut :

$$Entropy(S, jawaban_b \leq 10) = -\frac{50}{66} \log_2 \left(\frac{50}{66}\right) - \frac{16}{66} \log_2 \left(\frac{16}{66}\right) = 0,799$$

$$Entropy(S, jawaban_b > 10) = -\frac{13}{37} \log_2 \left(\frac{13}{37}\right) - \frac{23}{37} \log_2 \left(\frac{23}{37}\right) - \frac{1}{37} \log_2 \left(\frac{1}{37}\right) = 1,097$$

5) Menghitung nilai *entropy* jawaban\_b ≤20 dan >20

Atribut jawaban\_b ≤20 dan >20 memiliki 2 buah jenis *value* yaitu ≤20 sebanyak 102 buah dan >20 sebanyak 1 buah. Dengan demikian perhitungan nilai *entropy* untuk setiap *value* atribut diperoleh sebagai berikut :

$$Entropy(S, jawaban_b \leq 20) = -\frac{62}{66} \log_2 \left(\frac{62}{66}\right) - \frac{22}{66} \log_2 \left(\frac{22}{66}\right) - \frac{17}{66} \log_2 \left(\frac{17}{66}\right) = 1,337$$

$$Entropy(S, jawaban_b > 20) = -\frac{1}{1} \log_2 \left(\frac{1}{1}\right) = 0$$

6) Menghitung nilai *entropy* jawaban\_b ≤30 dan >30

Atribut jawaban\_b ≤30 dan >30 memiliki 2 buah jenis *value* yaitu ≤30 sebanyak 103 buah dan >30 sebanyak 0 buah. Dengan demikian perhitungan nilai *entropy* untuk setiap *value* atribut diperoleh sebagai berikut :

$$Entropy(S, jawaban_b \leq 30) = -\frac{63}{103} \log_2 \left(\frac{63}{103}\right) - \frac{23}{103} \log_2 \left(\frac{23}{103}\right) - \frac{17}{103} \log_2 \left(\frac{17}{103}\right) = 1,346$$

$$Entropy(S, jawaban_b > 30) = -\frac{0}{0} \log_2 \left(\frac{0}{0}\right) = 0$$

7) Menghitung nilai *entropy* jawaban\_c ≤10 dan >10

Atribut jawaban\_c ≤10 dan >10 memiliki 2 buah jenis *value* yaitu ≤10 sebanyak 83 buah dan >10 sebanyak 20 buah. Dengan demikian perhitungan nilai *entropy* untuk setiap *value* atribut diperoleh sebagai berikut :

$$Entropy(S, jawaban_c \leq 10) = -\frac{62}{83} \log_2 \left(\frac{62}{83}\right) - \frac{21}{83} \log_2 \left(\frac{21}{83}\right) = 0,819$$

$$Entropy(S, jawaban_c > 10) = -\frac{1}{20} \log_2 \left(\frac{1}{20}\right) - \frac{2}{20} \log_2 \left(\frac{2}{20}\right) - \frac{17}{20} \log_2 \left(\frac{17}{20}\right) = 0,748$$

8) Menghitung nilai *entropy* jawaban\_c ≤20 dan >20

Atribut jawaban\_c ≤20 dan >20 memiliki 2 buah jenis *value* yaitu ≤20 sebanyak 102 buah dan >20 sebanyak 1 buah. Dengan demikian perhitungan nilai *entropy* untuk setiap *value* atribut diperoleh sebagai berikut :

$$Entropy(S, jawaban_c \leq 20) = -\frac{63}{102} \log_2\left(\frac{63}{102}\right) - \frac{23}{102} \log_2\left(\frac{23}{102}\right) - \frac{17}{102} \log_2\left(\frac{17}{102}\right) = 1,333$$

$$Entropy(S, jawaban_c > 20) = -\frac{1}{1} \log_2\left(\frac{1}{1}\right) = 0$$

9) Menghitung nilai *entropy* jawaban\_c ≤30 dan >30

Atribut jawaban\_c ≤30 dan >30 memiliki 2 buah jenis *value* yaitu ≤30 sebanyak 103 buah dan >30 sebanyak 0 buah. Dengan demikian perhitungan nilai *entropy* untuk setiap *value* atribut diperoleh sebagai berikut :

$$Entropy(S, jawaban_c \leq 30) = -\frac{63}{103} \log_2\left(\frac{63}{103}\right) - \frac{23}{103} \log_2\left(\frac{23}{103}\right) - \frac{17}{103} \log_2\left(\frac{17}{103}\right) = 1,346$$

$$Entropy(S, jawaban_c > 30) = -\frac{0}{0} \log_2\left(\frac{0}{0}\right) = 0$$

d. Menghitung nilai *information gain* tiap atribut

$$Information\ Gain(S, F_j) = Entropy(S) - \sum_{v_i \in V_{F_j}} \frac{|S_{v_i}|}{|S|} \cdot Entropy(S_{v_i})$$

$$Information\ Gain(Jawaban_a \leq 10\ dan > 10) = 1,346 - \frac{27}{103} \cdot 0,999 - \frac{76}{103} \cdot 0,793 = 0,498$$

$$Information\ Gain(Jawaban_a \leq 20\ dan > 20) = 1,346 - \frac{91}{103} \cdot 1,422 - \frac{91}{103} \cdot 0 = 0,090$$

$$Information\ Gain(Jawaban_a \leq 30\ dan > 30) = 1,346 - \frac{103}{103} \cdot 1,346 = 0$$

$$Information\ Gain(Jawaban_b \leq 10\ dan > 10) = 1,346 - \frac{66}{103} \cdot 0,799 - \frac{37}{103} \cdot 1,097 = 0,440$$

$$Information\ Gain(Jawaban_b \leq 20\ dan > 20) = 1,346 - \frac{102}{103} \cdot 1,337 - \frac{1}{103} \cdot 0 = 0,021$$

$$Information\ Gain(Jawaban_b \leq 30\ dan > 30) = 1,346 - \frac{103}{103} \cdot 1,346 = 0$$

$$Information\ Gain(Jawaban_c \leq 10\ dan > 10) = 1,346 - \frac{83}{103} \cdot 0,816 - \frac{20}{103} \cdot 0,748 = 0,543$$

$$Information\ Gain(Jawaban_c \leq 20\ dan > 20) = 1,346 - \frac{102}{103} \cdot 1,333 - \frac{1}{103} \cdot 0 = 0,026$$

$$Information\ Gain(Jawaban_c \leq 30\ dan > 30) = 1,346 - \frac{103}{103} \cdot 1,346 = 0$$

e. Menentukan *root tree*

Informasi *gain* masing-masing atribut hasil perhitungan pertama seperti di bawah ini :

TABEL III. INFORMATION GAIN PERTAMA

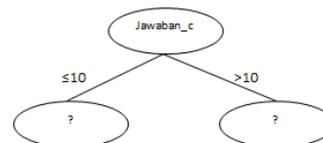
Atribut	Information Gain
Jawaban_a ≤10 dan >10	0,498
Jawaban_a ≤20 dan >20	0,090
Jawaban_a ≤30 dan >30	0
Jawaban_b ≤10 dan >10	0,440
Jawaban_b ≤20 dan >20	0,021
Jawaban_b ≤30 dan >30	0
Jawaban_c ≤10 dan >10	0,543
Jawaban_c ≤20 dan >20	0,026
Jawaban_c ≤30 dan >30	0

Berdasarkan perhitungan informasi *gain*, dapat dilihat bahwa nilai *gain* tertinggi dimiliki oleh atribut jawaban\_c ≤10 dan >10, sehingga dengan demikian jawaban\_c

dijadikan sebagai root dari pohon keputusan yang akan dibentuk.

Sedangkan simpul keputusan dilihat berdasarkan *value* atribut dari jawaban\_c, yakni jika *value* atribut memiliki class target berbeda, maka dilakukan perhitungan nilai *gain* informasi kembali terhadap atribut tertentu yang merujuk pada *value* atribut yang menjadi root pohon keputusan.

Berikut merupakan ilustrasi terbentuknya root pohon keputusan dengan simpul keputusan awal menurut *value* atribut yang dimilikinya.



Gambar 1. Root Pohon Keputusan Kesatu

f. Setelah atribut jawaban\_c ≤10 dan >10 menjadi atribut root, maka data *training* awal kemudian menyusut setelah terjadi split data *training*. Berikut informasi *gain* masing-masing atribut hasil perhitungan kedua seperti di bawah ini :

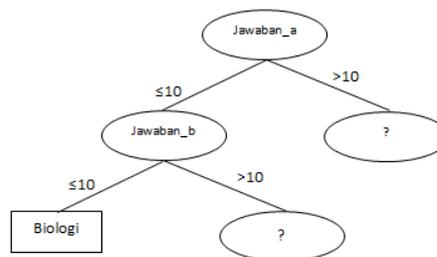
TABEL IV. INFORMATION GAIN KEDUA

Atribut	Information Gain
Jawaban_a ≤10 dan >10	0,312
Jawaban_a ≤20 dan >20	0,189
Jawaban_a ≤30 dan >30	0
Jawaban_b ≤10 dan >10	0,423
Jawaban_b ≤20 dan >20	0,024
Jawaban_b ≤30 dan >30	0

Berdasarkan perhitungan informasi *gain*, dapat dilihat bahwa nilai *gain* tertinggi dimiliki oleh atribut jawaban\_b ≤10 dan >10, sehingga dengan demikian jawaban\_b ≤10 dan >10 dijadikan sebagai root dari pohon keputusan yang akan dibentuk.

Sedangkan simpul keputusan dilihat berdasarkan *value* atribut dari jawaban\_b ≤10 dan >10, yakni jika *value* atribut memiliki class target berbeda, maka dilakukan perhitungan nilai *gain* informasi kembali terhadap atribut tertentu yang merujuk pada *value* atribut yang menjadi root pohon keputusan.

Berikut merupakan ilustrasi terbentuknya root pohon keputusan dengan simpul keputusan awal menurut *value* atribut yang dimilikinya.



Gambar 2. Leaf Pohon Keputusan Kedua

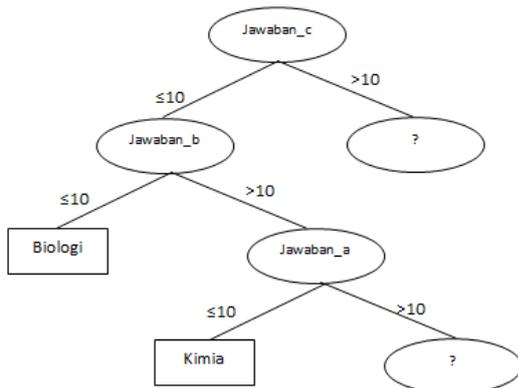
g. Setelah atribut jawaban\_b ≤10 menjadi atribut root, maka data training awal kemudian menyusut setelah terjadi split data training. Selanjutnya menentukan leaf node untuk jawaban\_b >10, Berikut informasi gain masing-masing atribut hasil perhitungan ketiga seperti di bawah ini :

TABEL V. INFORMATION GAIN KETIGA

Atribut	Information Gain
Jawaban_a ≤10 dan >10	0,292
Jawaban_a ≤20 dan >20	0
Jawaban_a ≤30 dan >30	0

Berdasarkan perhitungan informasi gain, dapat dilihat bahwa nilai gain tertinggi dimiliki oleh atribut jawaban\_a ≤10 dan >10, sehingga dengan demikian jawaban\_a ≤10 dan >10 dijadikan sebagai root dari pohon keputusan yang akan dibentuk. Sedangkan simpul keputusan dilihat berdasarkan value atribut dari jawaban\_a ≤10 dan >10, yakni jika value atribut memiliki class target berbeda, maka dilakukan perhitungan nilai gain informasi kembali terhadap atribut tertentu yang merujuk pada value atribut yang menjadi root pohon keputusan.

Berikut merupakan ilustrasi terbentuknya root pohon keputusan dengan simpul keputusan awal menurut value atribut yang dimilikinya.



Gambar 3. Leaf Pohon Keputusan Ketiga

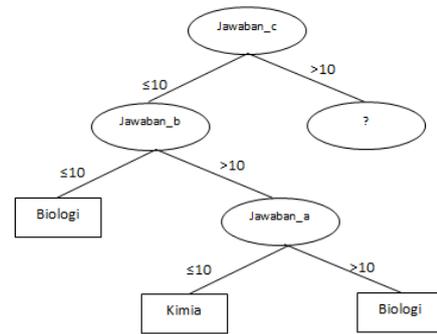
h. Setelah atribut jawaban\_a ≤10 menjadi atribut root, maka data training awal kemudian menyusut setelah terjadi split data training. Berikut informasi gain masing-masing atribut hasil perhitungan keempat seperti di bawah ini :

TABEL VI. INFORMATION GAIN KEEMPAT

Atribut	Information Gain
Jawaban_a ≤10 dan >10	0
Jawaban_a ≤20 dan >20	0
Jawaban_a ≤30 dan >30	0

Berdasarkan perhitungan informasi gain, dapat dilihat bahwa semua nilai gain atribut bernilai 0 sehingga perhitungan terhenti dan menghasilkan suatu keputusan.

Berikut merupakan ilustrasi terbentuknya root pohon keputusan dengan simpul keputusan awal menurut value atribut yang dimilikinya.



Gambar 4. Leaf Pohon Keputusan Keempat

i. Setelah pada langkah sebelumnya dilakukan penentuan atribut root dan terpilih jawaban\_b menjadi leaf node, maka data training awal kemudian menyusut setelah terjadi split data training. Berikut informasi gain masing-masing atribut hasil perhitungan kelima seperti di bawah ini :

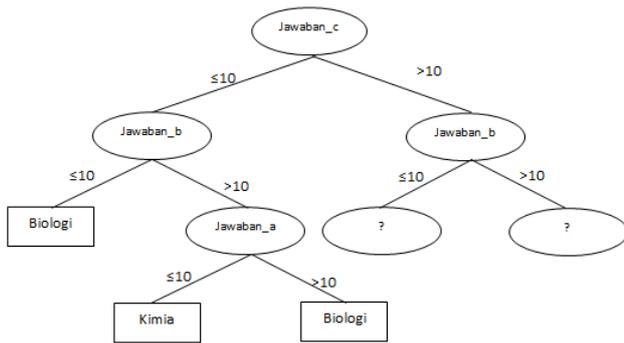
TABEL VII. INFORMATION GAIN KELIMA

Atribut	Information Gain
Jawaban_a ≤10 dan >10	0,138
Jawaban_a ≤20 dan >20	0
Jawaban_a ≤30 dan >30	0
Jawaban_b ≤10 dan >10	0,335
Jawaban_b ≤20 dan >20	0
Jawaban_b ≤30 dan >30	0

Berdasarkan perhitungan informasi gain, dapat dilihat bahwa nilai gain tertinggi dimiliki oleh atribut jawaban\_b ≤10 dan >10, sehingga dengan demikian jawaban\_b ≤10 dan >10 dijadikan sebagai root dari pohon keputusan yang akan dibentuk.

Sedangkan simpul keputusan dilihat berdasarkan value atribut dari jawaban\_b ≤10 dan >10, yakni jika value atribut memiliki class target berbeda, maka dilakukan perhitungan nilai gain informasi kembali terhadap atribut tertentu yang merujuk pada value atribut yang menjadi root pohon keputusan.

Berikut merupakan ilustrasi terbentuknya root pohon keputusan dengan simpul keputusan awal menurut value atribut yang dimilikinya.



Gambar 5. Leaf Pohon Keputusan Kelima

j. Setelah atribut jawaban\_b >10 menjadi atribut root, maka data training awal kemudian menyusut setelah terjadi split data training. Berikut informasi gain masing-masing atribut hasil perhitungan keenam seperti di bawah ini :

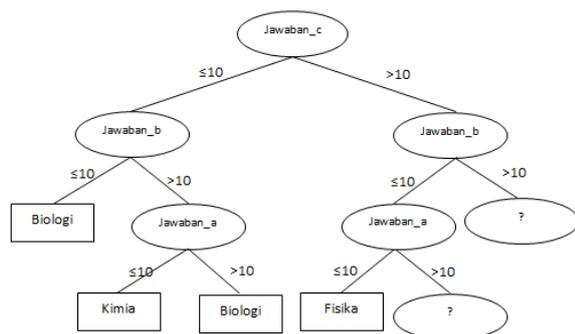
TABEL VII. INFORMATION GAIN KEENAM

Atribut	Information Gain
Jawaban a ≤10 dan >10	0,132
Jawaban a ≤20 dan >20	0
Jawaban a ≤30 dan >30	0

Berdasarkan perhitungan informasi gain, dapat dilihat bahwa nilai gain tertinggi dimiliki oleh atribut jawaban\_a ≤10 dan >10, sehingga dengan demikian jawaban\_a ≤10 dan >10 dijadikan sebagai root dari pohon keputusan yang akan dibentuk.

Sedangkan simpul keputusan dilihat berdasarkan value atribut dari jawaban\_a ≤10 dan >10, yakni jika value atribut memiliki class target berbeda, maka dilakukan perhitungan nilai gain informasi kembali terhadap atribut tertentu yang merujuk pada value atribut yang menjadi root pohon keputusan.

Berikut merupakan ilustrasi terbentuknya root pohon keputusan dengan simpul keputusan awal menurut value atribut yang dimilikinya.



Gambar 6. Leaf Pohon Keputusan Keenam

k. Setelah atribut jawaban\_a >10 menjadi atribut root, maka data training awal kemudian menyusut setelah terjadi split

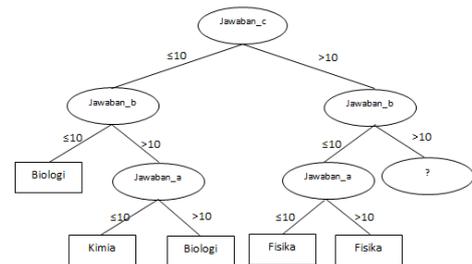
data training. Berikut informasi gain masing-masing atribut hasil perhitungan ketujuh seperti di bawah ini :

TABEL IX. INFORMATION GAIN KETUJUH

Atribut	Information Gain
Jawaban a ≤10 dan >10	0
Jawaban a ≤20 dan >20	0
Jawaban a ≤30 dan >30	0

Berdasarkan perhitungan informasi gain, dapat dilihat bahwa semua nilai gain atribut bernilai 0 sehingga perhitungan terhenti dan menghasilkan suatu keputusan.

Berikut merupakan ilustrasi terbentuknya root pohon keputusan dengan simpul keputusan awal menurut value atribut yang dimilikinya.



Gambar 7. Leaf Pohon Keputusan Ketujuh

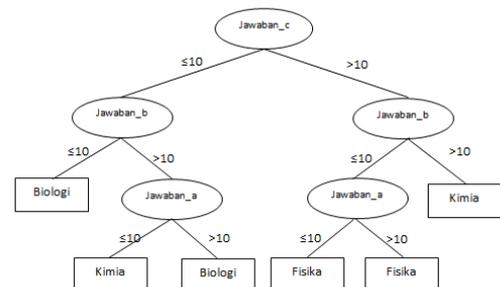
l. Setelah pada langkah sebelumnya dilakukan penentuan atribut root dan terpilih jawaban\_a menjadi leaf node, maka data training awal kemudian menyusut setelah terjadi split data training. Berikut informasi gain masing-masing atribut hasil perhitungan kedelapan seperti di bawah ini :

TABEL X. INFORMATION GAIN KEDELAPAN

Atribut	Information Gain
Jawaban a ≤10 dan >10	0
Jawaban a ≤20 dan >20	0
Jawaban a ≤30 dan >30	0

Berdasarkan perhitungan informasi gain, dapat dilihat bahwa semua nilai gain atribut bernilai 0 sehingga perhitungan terhenti dan menghasilkan suatu keputusan.

Berikut merupakan ilustrasi terbentuknya root pohon keputusan dengan simpul keputusan awal menurut value atribut yang dimilikinya.



Gambar 8. Leaf Pohon Keputusan Kedelapan

Berdasarkan pohon keputusan diatas maka diperoleh aturan rule sebagai berikut :

- 1) IF jawaban\_c ≤ 10 AND jawaban\_b ≤ 10 THEN Biologi
- 2) IF jawaban\_c ≤ 10 AND jawaban\_b > 10 AND jawaban\_a ≤ 10 THEN Kimia
- 3) IF jawaban\_c ≤ 10 AND jawaban\_b > 10 AND jawaban\_a > 10 THEN Biologi
- 4) IF jawaban\_c > 10 AND jawaban\_b ≤ 10 AND jawaban\_a ≤ 10 THEN Fisika
- 5) IF jawaban\_c > 10 AND jawaban\_b ≤ 10 AND jawaban\_a > 10 THEN Fisika
- 6) IF jawaban\_c > AND jawaban\_b > 10 THEN Kimia

Setelah mendapatkan aturan rule dari perhitungan data training maka selanjutnya adalah melakukan uji rule menggunakan data testing sebanyak 41 data sebagai berikut:

TABEL XI. UJI RULE DATA TESTING

No	Nama	Jawab A	Jawab B	Jawab C	Minat Asli	Rule	Minat Hasil	Tepat
1	Abi Saputra	5	12	13	Fisika	6	Kimia	Tidak Tepat
2	Amelia M	5	8	17	Fisika	4	Fisika	Tepat
3	Annisa Nuraini S	16	5	9	Biologi	1	Biologi	Tepat
4	Arya Putra Fauzi	8	8	14	Fisika	4	Fisika	Tepat
5	Bima Septaragi	12	6	12	Fisika	5	Fisika	Tepat
6	Elya Magpiroh	14	12	4	Biologi	3	Biologi	Tepat
7	Hanum Septiana	24	3	3	Biologi	1	Biologi	Tepat
8	Imas Masruroh	15	8	7	Biologi	1	Biologi	Tepat
9	Keny Yulianah	21	3	6	Biologi	1	Biologi	Tepat
10	Lena Apriyana	13	12	5	Biologi	3	Biologi	Tepat
11	Lisa Lidiana P	11	8	11	Fisika	5	Fisika	Tepat
12	Marsha Salisa A	15	8	7	Biologi	1	Biologi	Tepat
13	Mei Fuji R	11	12	7	Kimia	3	Biologi	Tidak Tepat
14	Mella Handayani	22	7	1	Biologi	1	Biologi	Tepat
15	Muhamad Gibran	9	15	6	Kimia	2	Kimia	Tepat
16	Mujriatul F	25	3	2	Biologi	1	Biologi	Tepat
17	Nabila Sri Taslim	30	0	0	Biologi	1	Biologi	Tepat
18	Nanda Angraeni	16	12	2	Biologi	3	Biologi	Tepat
19	Ninda Aulia	10	8	12	Fisika	4	Fisika	Tepat
20	Nurul Mufaridah	11	7	12	Fisika	5	Fisika	Tepat
21	Puput Melati	22	6	2	Biologi	1	Biologi	Tepat
22	Ravina Aulia	8	14	8	Kimia	2	Kimia	Tepat
23	Refalzi	9	9	12	Fisika	4	Fisika	Tepat
24	Resa Umami	26	3	1	Biologi	1	Biologi	Tepat
25	Rizal	8	7	15	Fisika	4	Fisika	Tepat
26	Septi Andriani	12	10	8	Biologi	1	Biologi	Tepat
27	Shalsabilla A. C	14	8	8	Biologi	1	Biologi	Tepat
28	Siti Khodijah	0	0	30	Fisika	4	Fisika	Tepat
29	Suherni	11	13	6	Kimia	3	Biologi	Tidak Tepat
30	Sulhah	19	5	6	Biologi	1	Biologi	Tepat
31	Surifatul Naima	10	10	10	Fisika	1	Biologi	Tidak Tepat
32	Sutihah	11	14	5	Kimia	3	Biologi	Tidak

No	Nama	Jawab A	Jawab B	Jawab C	Minat Asli	Rule	Minat Hasil	Tepat
33	Talitha Fauziah	16	8	6	Biologi	1	Biologi	Tepat
34	Tasya Adela T	6	10	14	Fisika	4	Fisika	Tepat
35	Teguh Rizki	11	11	8	Kimia	3	Biologi	Tidak Tepat
36	Tsabita Dayana	10	17	3	Kimia	6	Kimia	Tepat
37	Tusi Rahmawati	7	12	11	Kimia	2	Kimia	Tepat
38	Utari Ayu W	6	21	3	Kimia	2	Kimia	Tepat
39	Widya Ningsih	30	0	0	Biologi	1	Biologi	Tepat
40	Yulia Salsabila	11	13	6	Kimia	3	Biologi	Tidak Tepat
41	Zamelia Maulida	18	5	7	Biologi	1	Biologi	Tepat

### III. HASIL DAN PEMBAHASAN

Pada penelitian ini untuk menghitung kinerja dari klasifikasi ini menggunakan metode confusion matrix dari hasil uji rule data testing maka diperoleh nilai akurasi yaitu 82,82 %.

Setelah sistem selesai dirancang dan dibangun, evaluasi sistem dilakukan menggunakan kuesioner. Kuesioner diberikan kepada 10 responden di SMA Negeri 1 Anyer yang berbeda. Berikut tabel hasil kuesioner.

TABEL XII. REKAPITULASI HASIL KUESIONER

No	Nama Responden	P1	P2	P3	P4	P5
1	Mei Fuji Rahayu	4	4	5	5	4
2	Suherni	5	4	5	4	4
3	Ninda Aulia	4	5	5	5	4
4	Rizal	5	5	4	5	5
5	Annisa Nuraini S	4	4	4	4	4
6	Arya Putra Fauzi	5	4	4	4	4
7	Bima Septaragi	4	4	4	5	5
8	Faizal Bagus Z	5	4	4	5	5
9	Septi Andriani	4	5	5	4	5
10	Valentina	5	4	4	5	4

Berdasarkan rekapitulasi hasil kuesioner pada Tabel XII, maka selanjutnya dilakukan perhitungan hasil skor dengan menggunakan Skala Likert. Pengujian menggunakan skala likert merupakan skala yang paling banyak digunakan dalam riset berupa survey, dan pengujian penelitian deskriptif. Pada pengujian skala likert responden diminta untuk menentukan tingkat persetujuan mereka terhadap suatu pertanyaan dengan memilih salah satu pilihan yang telah tersedia [10].

Hasil skor kemudian dihitung persentase kelayakannya. Berikut merupakan hasil dari perhitungan persentase kelayakan setiap variabel yang disajikan dalam Tabel XIII.

TABEL XIII. HASIL PERHITUNGAN KUESIONER

No	Pernyataan	Keterangan					Total Skor	Interpretasi Index
		STS	TS	KS	S	SS		
1	Saudara/i membutuhkan adanya aplikasi klasifikasi kelas lintas minat	0	0	0	25	20	45	90 %
2	Saudara/i memerlukan informasi tentang mata pelajaran lintas minat	0	0	0	15	28	43	94 %
3	Saudara/i terbantu dengan adanya aplikasi lintas minat	0	0	0	20	24	44	88 %
4	Aplikasi lintas minat ini mudah digunakan oleh saudara/i	0	0	0	30	16	46	92 %

5	Saudara/i merasa puas ketika mengetahui hasil kelas lintas minat sesuai dengan anda	0	0	0	20	24	44	88 %
Rata-rata = Jumlah Total index % / Jumlah Pertanyaan								90,4 %

## VI. KESIMPULAN

Kesimpulan yang didapat pada penelitian ini adalah sebagai berikut :

- 1) Aplikasi sistem klasifikasi ini dapat memberikan hasil kelas lintas minat melalui atribut sebanyak 30 butir untuk setiap alternatifnya. Atribut yang digunakan diperoleh dari kriteria pada setiap mata pelajaran lintas minat.
- 2) Pada perhitungan ini terdapat 2 tahapan yaitu tahapan perhitungan data training dimana diambil sebanyak 103 data siswa untuk diolah. Dan tahapan pengujian data testing sebanyak 41 data siswa, pada tahapan ini didapat nilai akurasi sebesar 82,82%.
- 3) Model klasifikasi lintas minat siswa ini dapat digunakan sebagai alternatif referensi bagi guru BK untuk dapat mengelompokkan siswa berdasarkan minat dan bakat pada saat pohon keputusan ini terbentuk.

## UCAPAN TERIMA KASIH

Saya mengucapkan terima kasih kepada pihak-pihak yang telah membantu dalam penerbitan paper ini.

## REFERENSI

- [1] Permendikbud, Pedoman Peminatan Peserta Didik. Jakarta : Permendikbud, 2014.
- [2] P. Innike Marbitha, "Identifikasi Antusiasme Dan Hasil Belajar Siswa Kelas X IPS Pada Mata Pelajaran Biologi Program Peminatan Di SMA Negeri Colomadu Tahun Ajaran 2015/2016". Skripsi pada Universitas Muhammadiyah Surakarta. Tidak Diterbitkan, 2016.
- [3] P. Eko, Data Mining Mengolah Data Menjadi Informasi Menggunakan Matlab. Yogyakarta: Andi, 2014.
- [4] Whitten, Jeffrey L. dan Bentley, Lonnie D. (2010). Systems Analysis And Design Methods. New York: McGraw-Hill/Irwin.
- [5] Widodo, Prabowo P. dkk (2011). Pemodelan Sistem Berorientasi Obyek Dengan UML. Graha ilmu: Yogyakarta
- [6] Sugiyono, Metode Penelitian Kuantitatif Kualitatif dan R&D. Bandung: Alfabeta, 2009.
- [7] Turban, E. dkk, Decision Support Systems and Intelligent Systems. Yogyakarta: Andi, 2005.
- [8] Munandar, Tb Ai, Modul Perkuliahan Data mining Edisi Revisi 2. Serang: Universitas Serang Raya, 2018.
- [9] Joița, D, *Unsupervised static discretization methods in data mining*. Bucharest, Romania: Titu Maiorescu University, 2010.
- [10] Jogyanto, Metodologi Penelitian Sistem Informasi. Yogyakarta: Andi, 2008.