

PREDIKSI KELULUSAN PMPA DI SEKOLAH MENENGAH ATAS***THE ACCEPTANCE PREDICTION OF NEW SENIOR HIGH SCHOOL STUDENT BY PMPA LINE***Kania Azka Augustine¹, Andri²

Fakultas Ilmu Komputer, Program Studi Sistem Informasi

Universitas Bina Darma Palembang

Email: 161410135@student.binadarma.ac.id

Abstrak

Penelitian ini bertujuan untuk mengetahui prediksi PPDB SMP/MTs jalur PMPA berdasarkan prestasi akademik dan prestasi non akademik yang dimiliki peserta didik selama pendidikan SMP/MTs menggunakan *data mining* dengan metode *Decision Tree* dan *Algoritma C4.5*. Data yang digunakan dalam penelitian ini sebanyak 249 peserta didik dari tahun 2017 sampai 2019. Berdasarkan hasil analisis *data mining* menggunakan *RapidMiner* didapatkan hasil Total Nilai sebagai *root* (akar) dalam pembuatan *decision tree* dengan akurasi 89.19%.

Kata Kunci: *Data Mining, Klasifikasi, Decision Tree, Algoritma C4.5, PMPA, Kelulusan*

Abstract

This study aims to predict the acceptance of new Junior High School Students of PMPA line based on academic achievements and non-academic achievements of students during junior high school using data mining with the Decision Tree method and C4.5 Algorithm. The data used in this study were 249 students from 2017 to 2019. Based on the analysis of data mining using RapidMiner, the result obtained was Total Value as root in making decision trees with an accuracy of 89.19%.

Keywords: *Data Mining, Classification, Decision Tree, C4.5 Algorithm, PMPA.*

PENDAHULUAN

Teknologi saat ini berkembang pesat. Seiring dengan perkembangannya, berbagai aspek kehidupan manusia telah difasilitasi oleh kemajuan teknologi ini. Kemajuan teknologi membuat pekerjaan manusia yang semula dilakukan secara manual

menjadi lebih cepat, lebih sederhana, dan efisien. Dalam dunia pendidikan, perkembangan teknologi ini sangat membantu proses seleksi penerimaan peserta didik baru.

Penerimaan Peserta Didik Baru (PPDB) memiliki dua jalur, termasuk jalur Penelusuran Minat dan Prestasi

Akademik (PMPA) dan jalur ujian tertulis. Seleksi jalur PMPA dilakukan berdasarkan hasil prestasi akademik dan non-akademik dari calon peserta didik yang mendapatkan rekomendasi dari Kepala Sekolah dari mana peserta didik SMP / MTs dikirim.

Sekolah Menengah Atas (SMA) Negeri 8 Palembang menggunakan dua jalur untuk menerapkan PPDB, yaitu jalur Penelusuran Minat dan Prestasi Akademik (PMPA) dan jalur tes mandiri. Julukan sekolah unggulan, menjadi salah satu sekolah yang membuat peserta didik ingin mendaftarkan diri di SMA tersebut, khususnya lewat jalur PMPA. Jumlah peserta didik yang mendaftar tidak diimbangi dengan kebutuhan sekolah yang menyediakan jalur PMPA. Tahun 2017, total peserta didik yang mendaftar sebanyak 89 peserta didik. Tahun 2018, sebanyak 72 peserta didik. Sedangkan tahun 2019, sebanyak 88 peserta didik. Pada tahun 2017 sampai 2019, sekolah menerima peserta didik PMPA sebanyak 10% dari daya tampung keseluruhan peserta didik.

Hasil PMPA dapat diprediksi melalui beberapa hal, seperti melalui nilai peserta didik di SMP dari semester satu sampai semester lima, peringkat kelas, dan prestasi non akademik. Prediksi ini dapat dilakukan melalui *data mining*. *Data Mining* adalah bidang berbagai bidang ilmu dengan mengumpulkan teknik untuk

belajar dari pembelajaran mesin, pengenalan pola, statistik, basis data, dan masalah penanganan data [1]. Data PPDB jalur PMPA di SMA Negeri 8 Palembang dapat digunakan untuk menggali informasi dari *dataset* besar dengan menghasilkan informasi yang dapat digunakan untuk memprediksi kelulusan PMPA di SMA Negeri 8 Palembang.

Data mining terdiri dari tiga teknik, salah satunya adalah teknik klasifikasi. Klasifikasi adalah fungsi *data mining* yang menetapkan item dalam suatu kumpulan data untuk digolongkan ke dalam kategori atau kelas. Klasifikasi bertujuan memprediksi kelas target untuk setiap kasus dalam data [2]. Adapun metode *data mining* yang dilakukan untuk memprediksi kelulusan peserta didik yaitu metode klasifikasi *Decision Tree* (pohon keputusan) menggunakan *algoritma C4.5*.

Decision Tree (pohon keputusan) adalah suatu model prediksi dalam data mining yang menggunakan struktur berhirarki atau struktur pohon. Konsep pada pohon keputusan yaitu mengubah data menjadi pohon keputusan dan aturan-aturan keputusan. [3]. Sedangkan *algoritma C4.5* untuk membangun pohon keputusan dimulai dari pemilihan variabel sebagai akar, kemudian membuat cabang dari tiap-tiap nilai, membagi kasus dalam cabang dan mengulangi proses untuk setiap

cabang sampai semua kasus pada cabang tidak terdapat kelas yang sama [4].

Berdasarkan latar belakang tersebut, maka tujuan penelitian ini adalah untuk mengetahui variabel-variabel kelulusan peserta didik dengan menerapkan teknik *data mining* dalam memprediksi tingkat kelulusan peserta didik SMP/MTs jalur PMPA di SMA Negeri 8 Palembang.

LANDASAN TEORI

Data Mining

Data Mining merupakan ilmu yang memanfaatkan data yang sebelumnya kurang terpakai untuk mendapatkan suatu informasi atau pengetahuan baru [5].

Data Mining dikategorikan menjadi dua kategori utama, yaitu: [1].

1. *Descriptive Mining*, menemukan karakteristik data dalam basis data. Teknik yang termasuk *descriptive mining* adalah *clustering*, *association*, dan *sequential mining*.
2. *Predictive Mining*, menemukan pola dari data menggunakan variabel lain di masa depan. Teknik yang termasuk *predictive mining* adalah *classification*.

Berdasarkan definisi-definisi tentang *data mining* tersebut dapat disimpulkan bahwa *data mining* merupakan suatu proses pencarian secara otomatis pada suatu gudang data untuk menemukan pola atau model.

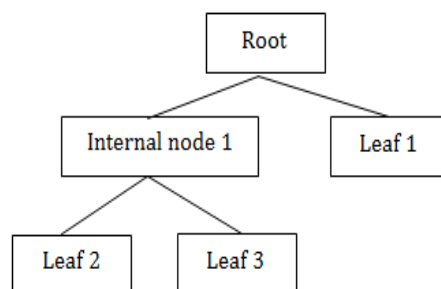
Klasifikasi

Klasifikasi adalah proses untuk menemukan model atau fungsi yang menjelaskan atau membedakan konsep atau kelas data, dengan tujuan untuk dapat memperkirakan kelas suatu objek yang labelnya tidak dapat diketahui [6].

Decision Tree (Pohon Keputusan)

Decision Tree (Pohon Keputusan) adalah model prediksi menggunakan struktur pohon atau struktur berhirarki. *Decision Tree* adalah *flow-chart* seperti struktur *tree*, dimana tiap cabang menunjukkan hasil dari *test* dan *leaf node* menunjukkan *class-class* atau *class distribution*.

Decision tree dapat mengelola nilai-nilai yang hilang atau data *noise*. Pada *decision tree* terdapat 3 jenis *node*, yaitu: [7]



Gambar 1. Struktur Susunan Pohon Keputusan (Andriani, 2013)

1. *Root node*, merupakan *node* paling atas, pada *node* ini tidak ada *input* dan bisa tidak mempunyai *output*

atau mempunyai *output* lebih dari satu.

2. *Internal node*, merupakan node percabangan, pada *node* ini hanya terdapat satu *input* dan mempunyai *output* minimal dua.
3. *Leaf node* atau *terminal node*, merupakan *node* akhir, pada *node* ini hanya terdapat satu *input* dan tidak mempunyai *output*.

Algoritma C4.5

Algoritma C4.5 dan pohon keputusan merupakan dua model yang tak terpisahkan, karena untuk membangun sebuah pohon keputusan, dibutuhkan *algoritma C4.5*. Pembuatan pohon keputusan menggunakan *algoritma C4.5* yang merupakan pengembangan dari *algoritma ID3*, dimana pengembangan dilakukan dalam hal mengatasi *missing data*, *data continue*, *pruning*. Secara umum, *algoritma C4.5* untuk membangun pohon keputusan dimulai dari pemilihan variabel sebagai akar, membuat cabang untuk tiap-tiap nilai, membagi kasus dalam cabang dan mengulangi proses untuk setiap cabang sampai semua kasus pada cabang memiliki kelas yang sama [8].

RapidMiner

RapidMiner adalah perangkat lunak yang bersifat *open source*. *RapidMiner* digunakan untuk analisis terhadap *data mining*, *text mining*, dan analisis prediksi. Dengan penggunaannya dapat memberikan

wawasan dalam membuat keputusan yang baik kepada pengguna menggunakan teknik *deskriptif* dan prediksi. *RapidMiner* mempunyai kurang lebih 500 operator *data mining*, termasuk *input*, *output*, *data preprocessing* dan visualisasi [9].

METODE PENELITIAN

Metode penelitian yang dilakukan sebagai berikut:

- a. Observasi
Teknik yang dilakukan dengan cara melakukan pengamatan secara langsung ke SMA Negeri 8 Palembang.
- b. Wawancara (*interview*)
Teknik yang dilakukan dengan cara melakukan sesi tanya jawab maupun wawancara secara langsung kepada wakil kesiswaan.
- c. Studi Pustaka
Mengumpulkan berbagai literatur dan referensi yang berkaitan dengan objek permasalahan.

Metode analisis data yang digunakan dalam penerapan *data mining* menggunakan metode tahapan *Knowledge Discovery in Databases (KDD)* yang terdiri dari: [9].

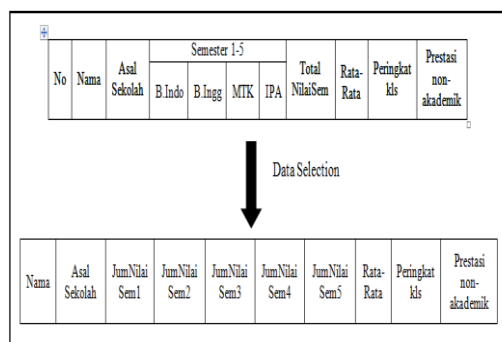
Data Selection

Pemilihan data harus dilakukan sebelum tahap pengumpulan data di *KDD* dimulai. Data dari hasil seleksi yang digunakan untuk *mining*

disimpan dalam file terpisah dari penyimpanan data yang dapat dieksekusi.

Data yang digunakan dalam penelitian ini adalah data peserta didik selama menempuh pendidikan SMP/MTs dari semester 1 sampai dengan semester 5 dari tahun 2017 sampai 2019 sebanyak 249 peserta didik. Semua variabel yang terdapat pada data peserta didik diantaranya nama, asal sekolah, nilai Bahasa Indonesia, Bahasa Inggris, Matematika, dan IPA dari semester 1 sampai semester 5, total nilai, rata-rata, dan prestasi non akademik. Tetapi variabel mata pelajaran Bahasa Indonesia, Bahasa Inggris, Matematika, dan IPA dari semester 1 sampai semester 5 tidak dipilih, karena variabel Total Nilai per semester peserta didik telah mewakili nilai mata pelajaran tersebut. Sedangkan variabel “Hasil” sebagai penentu Lulus dan Tidak Lulus peserta didik dalam pembentukan *decision tree* (pohon keputusan).

Proses *data selection* dalam penelitian ini dapat dilihat pada Gambar 2.



Gambar 2. Proses *Data Selection* (Agustine, 2020)

Preprocessing Data Cleaning

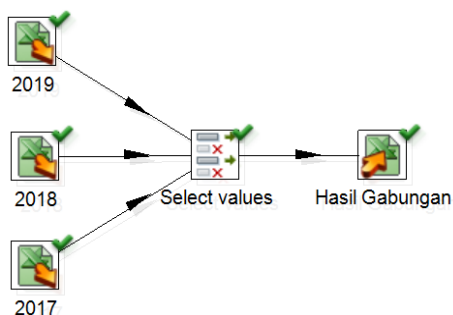
Seluruh variabel pada *dataset* tersebut akan diseleksi sedemikian rupa sehingga didapatkan variabel-variabel yang berisi nilai-nilai relevan. Syarat yang harus dilakukan dalam *preprocessing* adalah tidak terdapat *missing value* dan *redundant* sehingga menghasilkan sebuah *dataset* yang berisi dan siap digunakan untuk proses selanjutnya. *Missing value* adalah variabel-variabel yang tidak berisi nilai atau kosong dalam *dataset*, sementara *redundant* adalah jika dalam satu atau lebih *dataset* yang terdapat lebih dari satu *record* data yang berisi nilai ataupun data yang sama.

Data yang telah diperoleh dari SMA Negeri 8 Palembang masih terdapat bagian yang kosong yaitu bagian kolom hasil kelulusan peserta didik yaitu keterangan lulus/tidak lulus. Hal ini karena pihak sekolah hanya membatasi kelulusan peserta didik berdasarkan kuota yang disediakan pihak sekolah, sehingga diurutkan berdasarkan dari nilai tertinggi hingga nilai terendah, serta prestasi non akademik yang diperoleh apabila total nilainya tinggi.

Data Integration

Tahap integrasi data adalah tahap penggabungan data dari berbagai sumber. Adapun data yang digunakan dalam penelitian ini adalah data peserta didik dari tahun 2017 sampai 2019, lalu digabungkan menjadi satu untuk keperluan dalam proses *data mining*.

Proses integrasi data menggunakan aplikasi *Pentaho Data Integration*. Tahap yang dilakukan untuk proses integrasi tidak terdapat *step select value*, karena *step select value* telah dilakukan di tahap *data selection*. Dalam proses integrasi ini terdapat tambahan *step Microsoft Excel Input* sebanyak tiga step, karena terdapat tiga excel, yaitu tahun 2017, tahun 2018, dan tahun 2019, prosesnya dapat dilihat pada Gambar 3.



Gambar 3. Proses *Data Integration*
(Agustine, 2020)

Data Transformation

Transformation adalah proses pengubahan data yang dipilih, sehingga data tersebut cocok untuk proses *data mining*. Data yang akan digunakan dalam proses *data mining* formatnya belum bisa langsung digunakan, maka dari itu perlu adanya perubahan format agar bisa digunakan.

Dalam variabel prestasi non akademik yang diperoleh oleh peserta didik berbeda-beda, seperti taekwondo, tari, karate, dll. Maka dari itu, dalam tabel prestasi non akademik akan berubah menjadi “ada” (jika peserta didik tersebut memiliki prestasi) yang akan mempermudah dalam proses *data mining*.

HASIL DAN PEMBAHASAN

Data Mining

Data Mining adalah proses ekstraksi pola atau informasi menarik dari *dataset* besar yang dipilih menggunakan teknik tertentu. Dalam penelitian ini, teknik *data mining* yang digunakan adalah *classification* menggunakan *algoritma C4.5* yang digunakan sebagai variabel penentu tingkat kelulusan PMPA peserta didik. Penggunaan *algoritma C4.5* berperan penting dalam pembuatan pohon keputusan (*decision tree*). Proses yang dilakukan dalam analisis *data mining* menggunakan aplikasi *RapidMiner*.

Tahap yang dilakukan untuk menghitung *algoritma C4.5*, adalah sebagai berikut:

1. Menyiapkan *data training*. Data yang akan dilakukan dalam proses *data mining* sebanyak 249 peserta didik yang sebelumnya sudah dikelompokkan ke dalam kelas tertentu.

2. Menentukan *root* (akar), dengan cara menghitung *entropy* dari setiap variabel. Kemudian menghitung nilai *gain*, nilai *gain* yang tertinggi akan menjadi *root* dalam pembentukan pohon keputusan.

Adapun rumus entropi, yaitu:

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

Ket: S= himpunan kasus

A= variabel

n= jumlah partisi S

p_i = proporsi dari S_i kepada S

3. Menghitung nilai *gain*
Adapun rumus menghitung nilai *gain*, yaitu:

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

Ket: S= himpunan kasus

A= variabel

n= jumlah partisi variabel A

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

$|S|$ = jumlah kasus dalam S

Contoh perhitungan nilai *entropy* dan *gain* berdasarkan data sebanyak 249 peserta didik dapat dilihat pada Tabel 1.

Tabel 1. Contoh Perhitungan *Entropy* dan *Gain*

Node		Jumlah Data	Lulus	Tidak Lulus	Entropy	Gain
Total		249	108	141	0.987	
Sekolah						0.030
	MTs	2	2	0	0	
	SMP Negeri	228	103	125	0.993	
	SMP Swasta	19	3	16	0.629	
Total Nilai						
	>174	92	87	5	0.305	0.579
	≤174	157	21	136	0.469	
Rank Sem 1						0.387
	>2	129	13	116	0.987	
Total	≤2	120	95	25		

(Agustine, 2020)

Berikut ini contoh perhitungan variabel Total Nilai menggunakan rumus *entropy*, sebagai berikut:

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

$$= \left(-\frac{108}{249} \times \log_2 \left(\frac{108}{249} \right) \right) + \left(-\frac{141}{249} \times \log_2 \left(\frac{141}{249} \right) \right)$$

$$= 0.987$$

Sedangkan untuk menghitung *gain* sebagai berikut:

$$\begin{aligned}
Gain(S, A) &= Entropy(S) \\
&\quad - \sum_{i=1}^n \frac{|s_i|}{|S|} \times Entropy \\
&= 0.987292732 - \left(\left(\frac{92}{249} \right) \times 0.30 \right) \\
&\quad + \left(\left(\frac{157}{249} \right) \times 0.47 \right) \\
&= 0.579
\end{aligned}$$

Hasil penelitian tersebut terdiri dari variabel Total Nilai yang memiliki nilai *gain* tertinggi dipilih sebagai *root* (akar) dalam pembuatan *decision tree* dan merupakan kriteria yang paling menentukan hasil prediksi kelulusan peserta didik. Secara berurutan simpul berikutnya diperoleh oleh variabel-variabel yang bernilai *gain* lebih rendah dibandingkan dengan *root*, Seperti Rank Sem dan Asal Sekolah. Kemudian proses tersebut akan berhenti pada simpul akhir yang menyajikan hasil akhir Lulus dan Tidak Lulus dari setiap cabang yang disebut *leaf* (daun) sehingga membentuk menjadi pohon keputusan.

Implementasi RapidMiner

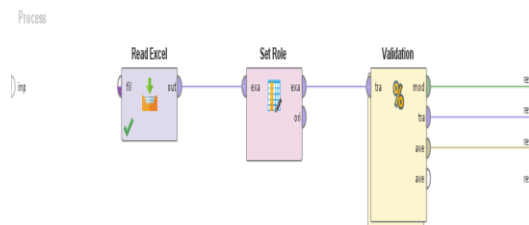
RapidMiner memiliki operator-operator yang akan digunakan dalam proses data mining. Adapun operator-operator yang akan dilakukan dalam penelitian ini, antara lain: [9]

1. *Read Excel*, proses ini dilakukan dengan mengambil data *excel* yang tabelnya telah digabung menggunakan *Pentaho Data*

Integration. Tipe data pada *excel* yang telah dipilih otomatis akan ditentukan oleh aplikasi.

2. *Set Role*, proses ini dilakukan untuk mengubah peran suatu variabel. Dalam penelitian ini mengubah variabel “Hasil” menjadi *label* yang akan menghasilkan “Lulus” dan “Tidak Lulus” sebagai hasil akhir bentuk pohon keputusan.

Validation (Split Validation), proses ini dilakukan untuk memperkirakan seberapa akurat suatu model dalam *data mining* yang prosesnya dapat dilihat pada Gambar 4.

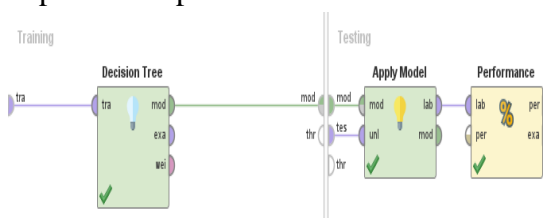


Gambar 4. Proses *Data Mining* menggunakan *Validation* (Agustine, 2020)

Operator *validation* memiliki dua subproses, yaitu *training* dan *testing*. Subproses *training* digunakan untuk melatih suatu model. Proses *training* digunakan untuk memilih metode *data mining*, seperti dalam penelitian ini menggunakan metode *Decision Tree*, maka dalam subproses *training* terdapat operator *Decision Tree*.

3. *Decision Tree* (Pohon Keputusan), proses ini dilakukan sesuai metode yang digunakan dalam penelitian.
4. *Apply Model*, tujuannya untuk memperoleh prediksi pada *data training* pada *unlabeled data (data testing)*. Serta sebagai penghubung antara *Decision Tree* dengan *Performance*. Proses ini yang paling penting adalah harus memiliki kesamaan urutan, variabel, dll antara *data testing* dan *data training*.
5. *Performance*, proses ini dilakukan untuk menghitung keakuratan suatu model.

Model subproses dari *validation (split validation)* yaitu proses *decision tree* di *training*, serta *apply model*, dan *performance* di *testing* yang prosesnya dapat dilihat pada Gambar 5.



Gambar 5. Model Subproses
Validation Decision Tree
(Agustine, 2020)

Berdasarkan *confusion matrix performance*, terdapat *accuracy*, *precision*, dan *recall*. Pembentukan hasil subproses dari *validation decision tree* telah dilakukan, selanjutnya mengetahui *accuracy*, *precision*, dll dari *Performance Vector*.

1. *Accuracy*, rasio yang menghasilkan prediksi *True (Positive dan Negative)* berdasarkan keseluruhan data. “Berapa persen peserta didik yang benar diprediksi Lulus dan Tidak Lulus dari keseluruhan peserta didik? “. Hasil *accuracy* pada penelitian ini adalah 89.19%, dapat dilihat pada Gambar 6.
2. *Precision*, rasio yang menghasilkan prediksi *True (Positive)* dibandingkan keseluruhan hasil yang diprediksi *Positive*. “Berapa persen peserta didik yang benar Tidak Lulus dari keseluruhan peserta didik yang diprediksi Lulus dan Tidak Lulus? “. Hasil *precision* pada penelitian ini adalah 92.50%, dapat dilihat pada Gambar 7.
3. *Recall*, rasio yang menghasilkan prediksi *True (Positive)* dibandingkan keseluruhan data yang sebenarnya diprediksi *Positive*. “Berapa persen peserta didik yang diprediksi Lulus dan Tidak Lulus dibandingkan keseluruhan peserta didik yang sebenarnya Tidak Lulus?“. Hasil *recall* pada penelitian ini adalah 88.10%, dapat dilihat pada Gambar 8.

Table View Plot View

accuracy: 88.19%

	true LULUS	true TIDAK LULUS	class precision
pred. LULUS	29	5	85.29%
pred. TIDAK LULUS	3	37	92.50%
class recall	90.62%	88.10%	

Gambar 6. Accuracy Performance Decision Tree (Agustine, 2020)

Table View Plot View

precision: 92.50% (positive class: TIDAK LULUS)

	true LULUS	true TIDAK LULUS	class precision
pred. LULUS	29	5	85.29%
pred. TIDAK LULUS	3	37	92.50%
class recall	90.62%	88.10%	

Gambar 7. Precision Performance Decision Tree (Agustine, 2020)

Table View Plot View

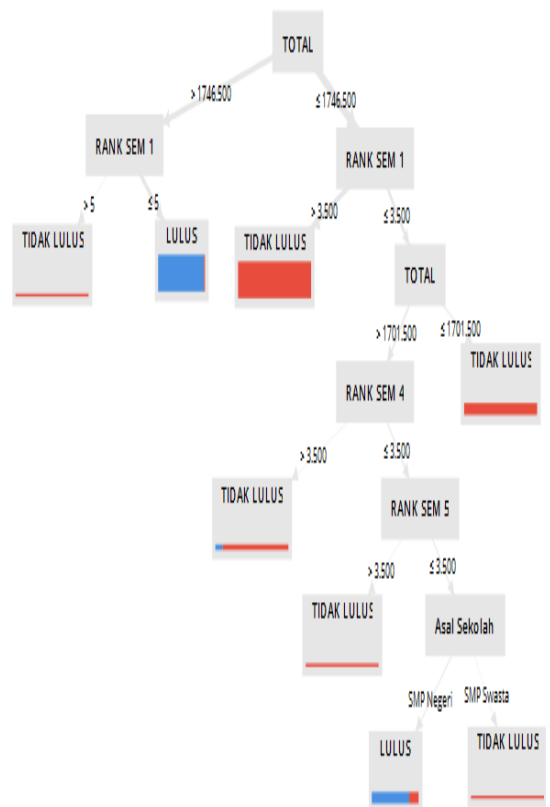
recall: 88.10% (positive class: TIDAK LULUS)

	true LULUS	true TIDAK LULUS	class precision
pred. LULUS	29	5	85.29%
pred. TIDAK LULUS	3	37	92.50%
class recall	90.62%	88.10%	

Gambar 8. Recall Performance Decision Tree

Dalam pohon keputusan tersebut pada Gambar 9 yang telah terbentuk, Jumlah Nilai Sem 1 sampai Sem 5, Prestasi Non Akademik dan Tingkat tidak termasuk kedalam kriteria faktor kelulusan peserta didik berdasarkan *algoritma C4.5*. Maka variabel tersebut *pruned* (dipangkas) dan tidak terpilih sebagai atribut pembentukan *decision tree*. Sedangkan variabel Total Nilai, Rank Sem, dan Asal Sekolah sebagai variabel dalam pohon keputusan.

Jumlah Nilai Sem 1 sampai Sem 5, Prestasi Non Akademik dan Tingkat tidak termasuk kedalam kriteria faktor kelulusan peserta didik berdasarkan *algoritma C4.5*. Maka variabel tersebut *pruned* (dipangkas) dan tidak terpilih sebagai atribut pembentukan *decision tree*. Sedangkan variabel Total Nilai, Rank Sem, dan Asal Sekolah sebagai variabel dalam pohon keputusan.



Gambar 9. Pohon Keputusan (Agustine, 2020)

Maka dari itu, dapat disimpulkan bahwa tidak seluruh variabel yang berpengaruh dalam kelulusan peserta didik, hanya dibutuhkan beberapa variabel yang berperan dalam pembentukan pohon keputusan. Kemudian terbentuklah daftar aturan dalam pada pohon keputusan tersebut. Daftar aturan tersebut dapat dilihat pada Gambar 10.

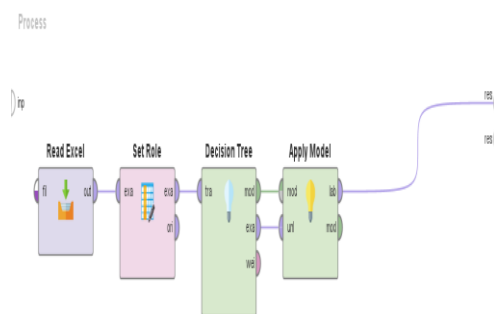
Tree

```
TOTAL > 1746.500
| RANK SEM 1 > 5: TIDAK LULUS (LULUS=0, TIDAK LULUS=3)
| RANK SEM 1 ≤ 5: LULUS (LULUS=97, TIDAK LULUS=2)
TOTAL ≤ 1746.500
| RANK SEM 1 > 3.500: TIDAK LULUS (LULUS=0, TIDAK LULUS=89)
| RANK SEM 1 ≤ 3.500
| | TOTAL > 1701.500
| | | RANK SEM 4 > 3.500: TIDAK LULUS (LULUS=1, TIDAK LULUS=9)
| | | RANK SEM 4 ≤ 3.500
| | | | RANK SEM 5 > 3.500: TIDAK LULUS (LULUS=0, TIDAK LULUS=4)
| | | | RANK SEM 5 ≤ 3.500
| | | | | Asal Sekolah = SMP Negeri: LULUS (LULUS=20, TIDAK LULUS=5)
| | | | | Asal Sekolah = SMP Swasta: TIDAK LULUS (LULUS=0, TIDAK LULUS=3)
| | | | | TOTAL ≤ 1701.500: TIDAK LULUS (LULUS=0, TIDAK LULUS=26)
```

Gambar 10. Daftar Aturan Pohon Keputusan (Agustine, 2020)

Setelah melakukan proses *decision tree* untuk mengetahui *accuracy*, *precision*, dll menggunakan *Split Validation*, kemudian mengetahui prediksi penggunaan metode *data mining* tersebut. Dalam memprediksi data kelulusan PMPA peserta didik, menggunakan operator *read excel* untuk membaca data excel, *set role* untuk menentukan label, *decision tree* untuk model *data mining*, serta *apply model* untuk melihat hasil prediksi

yang dilakukan di *RapidMiner* yang prosesnya dapat dilihat pada Gambar 11. Kemudian hasil kelulusan yang sesuai dan tidak sesuai pada data *excel* dengan hasil prediksi kelulusan dari operator *Apply Model* dapat dilihat pada Gambar 12 dan Gambar 13.



Gambar 11. Prediksi *Decision Tree* (Agustine, 2020)

Row No.	HASIL	prediction(HL)	confidence...	confidence...	Nama Siswa	Asal Sekolah	JUM
1	LULUS	LULUS	0.978	0.022	SYIFA SJAKL...	SMP Negeri	369
2	LULUS	LULUS	0.978	0.022	SEPTIANI MA...	SMP Swasta	376
3	LULUS	LULUS	0.978	0.022	M. RIDWAN F...	SMP Negeri	361
4	LULUS	LULUS	0.978	0.022	SYELLA NUR...	SMP Negeri	336
5	LULUS	LULUS	0.978	0.022	Aqila Amran ...	SMP Negeri	385
6	LULUS	LULUS	0.978	0.022	SEKAR ADEL...	SMP Negeri	361
7	LULUS	LULUS	0.978	0.022	ADELIA RAH...	SMP Negeri	371
8	LULUS	LULUS	0.978	0.022	ANNISA APRIL...	SMP Negeri	355
9	LULUS	LULUS	0.978	0.022	Fatimah Azza...	SMP Negeri	359
10	LULUS	LULUS	0.978	0.022	RIFDA DYAA A...	SMP Negeri	354

Gambar 12. Hasil Prediksi yang Sesuai (Agustine, 2020)

Row No.	Hasil	prediction(H...	confidence(...	confidence(...	Nama Siswa	Asal Seko
107	LULUS	LULUS	0.800	0.200	FAZETTI NAZ...	SMP Nege
108	LULUS	LULUS	0.800	0.200	NABILA ARIN...	SMP Nege
109	TIDAK LULUS	LULUS	0.978	0.022	SABRINA FIR...	SMP Nege
110	TIDAK LULUS	LULUS	0.978	0.022	AFIFA AMBAR	SMP Nege
111	TIDAK LULUS	TIDAK LULUS	0	1	REKA REGINA	SMP Nege
112	TIDAK LULUS	TIDAK LULUS	0	1	NIKEN ANGE...	SMP Nege
113	TIDAK LULUS	TIDAK LULUS	0	1	AZIZ ALKATOMI	SMP Nege
114	TIDAK LULUS	TIDAK LULUS	0	1	AHMAD FIKR...	SMP Nege
115	TIDAK LULUS	LULUS	0.800	0.200	Affiah Zahwa	SMP Nege
116	TIDAK LULUS	TIDAK LULUS	0.100	0.900	AZZURA AUR...	SMP Nege

Gambar 13. Hasil Prediksi yang Tidak Sesuai
(Agustine, 2020)

KESIMPULAN DAN SARAN

Berdasarkan hasil penelitian yang telah dilakukan dan yang telah diuraikan dalam Penerapan *Data Mining* untuk Prediksi PPDB SMP/MTs jalur PMPA di SMA Negeri 8 Palembang, maka peneliti dapat menyimpulkan bahwa variabel Jumlah Total Nilai adalah variabel yang memiliki *Gain* tertinggi dan menjadi *root* (akar) pada metode *Decision Tree* dan menjadi faktor utama dalam penentu kelulusan peserta didik SMP/MTs jalur PMPA di SMA Negeri 8 Palembang. Variabel Total Nilai, Rank Sem dan Asal Sekolah adalah variabel yang menentukan tingkat kelulusan PMPA di waktu yang akan datang. Dalam melakukan klasifikasi pada penelitian ini menggunakan metode *Decision Tree* dengan aplikasi *RapidMiner*, menghasilkan akurasi 89.19%, *precision* 92.50%, dan *recall* 88.10%. Hal ini menunjukkan bahwa *decision tree* memiliki performa yang handal dalam melakukan klasifikasi.

DAFTAR PUSTAKA

- [1] Yesi Novaria Kunang, S. Murniati, And Andri, "Implementasi Teknik Data Mining Untuk Memprediksi Tingkat Kelulusan Mahasiswa Pada Universitas Bina Darma Palembang," 2013, DOI: 10.13140/Rg.2.1.4212.1845.
- [2] B. Gupta, A. Rawat, A. Jain, A. Arora, And N. Dharmi, "Analysis Of Various Decision Tree Algorithms For Classification In Data Mining," *Int. J. Comput. Appl.*, Vol. 163, No. 8, Art. No. 8, Apr. 2017, DOI: 10.5120/Ijca2017913660.
- [3] S. Haryati, A. Sudarsono, And E. Suryana, "Implementasi Data Mining Untuk Memprediksi Masa Studi Mahasiswa Menggunakan Algoritma C4.5 (Studi Kasus: Universitas Dehasen Bengkulu)," Vol. 11, No. 2, Art. No. 2, 2015.
- [4] L. Elvitaria, "Memprediksi Tingkat Peminat Ekstrakurikuler Pada Siswa SMK Analisis Kesehatan Abdurrah Menggunakan Algoritma C4.5 (Studi Kasus: SMK Analisis Kesehatan Abdurrah)," *Rabit J. Teknol. Dan Sist. Inf. Univrab*, Vol. 2, No. 2, Art. No. 2, Aug. 2017, Doi: 10.36341/Rabit.V2i2.212.
- [5] Indrayanti, D. Sugianti, And M. A. Al Karomi, Eds., *Optimasi*

- Parameter K Pada Algoritma K-Nearest Neighbour Untuk Klasifikasi Penyakit Diabetes Mellitus.* 2017.
- [6] S. L. Br Ginting, W. Zarman, And A. Darmawan, “*Teknik Data Mining Untuk Memprediksi Masa Studi Mahasiswa.*” *Jurnal Teknik Komputer Unikom – Komputika*, 2014.
- [7] A. Andriani, “*Sistem Pendukung Keputusan Berbasis Decision Tree Dalam Pemberian Beasiswa Studi Kasus: AMIK ‘BSI Yogyakarta,*” P. 6, 2013.
- [8] R. T. Shita And N. Marliani, “*Aplikasi Data Mining Dengan Metode Classification Berbasis Algoritma C4.5.Pdf.*” *Seminar Nasional Sistem Informasi Indonesia*, 4 Desember 2013.
- [9] Rapidminer, *Rapidminer User Manual.* 2015.