

# Penerapan Educational Data Mining Untuk Memprediksi Hasil Belajar Siswa SMAK Ora et Labora

Daniel David<sup>a1</sup>, Sani M. Isa<sup>a2</sup>

<sup>a</sup> Department of Computer Science  
Bina Nusantara (BINUS) University  
Jakarta, Indonesia

<sup>1</sup> daniel.david001@binus.ac.id

<sup>2</sup> sani.m.isa@binus.ac.id

## Abstrak

*Data mining adalah salah satu alternatif yang bisa dilakukan untuk melakukan penggalian informasi baru dari sejumlah data yang besar. Salah satu aliran data mining adalah Educational Data Mining (EDM). EDM adalah aliran data mining yang bergerak pada bidang pendidikan. Dengan memanfaatkan data-data yang berhubungan dengan pendidikan, proses data mining bisa dilakukan untuk menemukan informasi berguna untuk kemajuan dalam bidang pendidikan. Penelitian ini menggunakan EDM dengan tujuan untuk memanfaatkan data internal assessment dari masing-masing siswa sekolah dan melakukan prediksi terhadap hasil ujian akhir nasional siswa tersebut. Data mining ini menggunakan teknik klasifikasi dan metode Decision Tree C4.5. Selain itu akan digunakan juga metode penelitian deskriptif agar bisa memberikan hasil yang lebih akurat. Penelitian ini diharapkan bisa memberikan kontribusi dalam bentuk prediksi hasil ujian akhir nasional sehingga kedepannya bisa digunakan untuk siswa angkatan seterusnya.*

**Kata Kunci:** Educational Data Mining, Klasifikasi, Decision Tree, Ujian Akhir Nasional

## 1. Pendahuluan

Menjelang Ujian Nasional, siswa SMA sering dibuat tidak tenang dalam belajar di sekolah karena tuntutan prestasi yang tinggi. Sedangkan hal ini kadang tidak sebanding dengan kemampuan belajar siswa tersebut. Oleh karena itu, banyak siswa yang mengalami stres akibat tidak bisa mengikuti pelajaran dengan baik ditambah dengan adanya tuntutan prestasi. Banyak faktor yang bisa berpengaruh dalam hasil belajar seorang siswa. Faktor-faktor seperti keluarga, kehidupan sosial, demografi siswa bisa berpengaruh dalam proses pembelajaran. Untuk itu diperlukan suatu informasi yang digali dari data siswa tersebut yang bisa digunakan untuk merubah strategi pembelajaran agar siswa nantinya bisa mendapatkan prestasi yang baik.

SMAK Ora et Labora mengalami hal-hal seperti banyaknya siswa yang mengalami nilai ujian akhir nasional berada dibawah rata-rata, baik rata-rata sekolah maupun rata-rata nasional. Bahkan pada tahun 2015, ada siswa yang dinyatakan tidak lulus dan harus mengulang dengan mengambil paket C. Hal ini menjadi perhatian khusus dalam penelitian kali ini. Dengan tujuan untuk melakukan prediksi terhadap siswa yang saat ini sedang berada di semester 4 berdasarkan data model yang diambil dari nilai siswa yang telah lulus, diharapkan sekolah nantinya bisa melakukan pencegahan terhadap siswa yang diprediksi akan mengalami nilai ujian akhir nasional berada di bawah rata-rata atau bahkan memiliki kemungkinan sampai tidak lulus.

Dalam penulisan paper penelitian ini, kami menyarankan penggunaan data mining untuk mengatasi permasalahan ini. Data mining dikenal sangat kuat dalam membongkar informasi tersembunyi dalam data dengan volume yang besar dan banyak digunakan dalam berbagai bidang salah satunya adalah pendidikan sehingga disebut dengan Educational Data Mining [1]. Ada beberapa teknik dan metode yang dapat digunakan dalam data mining. Teknik klasifikasi merupakan salah satu yang sering digunakan [1]. Metode Decision Tree C4.5 masuk ke dalam salah satu metode yang paling sering digunakan untuk data mining [2]. Banyak peneliti

menggunakan teknik ini karena kesederhanaannya dan keseluruhannya untuk membuka struktur data dalam skala kecil maupun besar dan memprediksi nilainya. Selain itu model dari Decision Tree mudah untuk dimengerti karena proses pemikirannya dan bisa dikonversikan secara langsung ke dalam IF-THEN rules [3].

Struktur dari paper ini terdiri dari 5 (lima) bagian. Bagian pertama memberikan pengenalan dan dasar-dasar penelitian ini. Bagian kedua mendiskusikan penelitian-penelitian sebelumnya yang berhubungan dengan bidang ini. Bagian ketiga menjelaskan metodologi yang akan digunakan dalam penelitian ini. Bagian keempat akan memberikan hasil, evaluasi, serta diskusi-diskusi. Dan bagian kelima akan menyimpulkan hasil keseluruhan dari penelitian ini.

## 2. Penelitian Terkait

Sebelumnya, beberapa penelitian telah dijalankan yang sesuai dengan bidang Educational Data Mining atau hanya sekedar data mining yang bergerak di bidang pendidikan. Penelitian-penelitian sebelumnya berjalan di berbagai macam aspek pendidikan yang diujikan. Aspek yang dijadikan penelitian berupa pemantauan performa dari siswa, prediksi terhadap karir siswa kedepannya, prediksi rekomendasi beasiswa atau prediksi masa studi siswa. Teknik yang digunakan kebanyakan adalah klasifikasi, dan metode yang digunakan adalah Decision Tree, SVM, Neural Network, dan Naive Bayes. Sedangkan atribut penelitian yang digunakan kebanyakan adalah internal assessment, demografi siswa, interaksi sosial, dan aktifitas ekstrakurikuler. Berikut adalah tabel dari beberapa penelitian terkait sebelumnya.

**Tabel 2.1** Studi Literatur

No.	Author	Teknik	Metode	Atribut
1.	[4]	<ul style="list-style-type: none"> <li>• Classification</li> <li>• Clustering</li> </ul>	<ul style="list-style-type: none"> <li>• Decision Tree</li> <li>• Rule Induction</li> <li>• Artificial NN</li> <li>• K-NN</li> <li>• Naïve Bayes</li> <li>• Random Forest</li> </ul>	<ul style="list-style-type: none"> <li>• Internal Assessment</li> </ul>
2.	[5]	<ul style="list-style-type: none"> <li>• Classification</li> <li>• Regression</li> </ul>	<ul style="list-style-type: none"> <li>• FFNN</li> <li>• SVM</li> <li>• PESFAM</li> <li>• LOGIT_Act</li> <li>• SEDM</li> </ul>	<ul style="list-style-type: none"> <li>• Internal Assessment</li> </ul>
3.	[6]	<ul style="list-style-type: none"> <li>• Clustering</li> <li>• Sequential Pattern</li> </ul>	<ul style="list-style-type: none"> <li>• K-means</li> </ul>	<ul style="list-style-type: none"> <li>• Internal Assessment</li> </ul>
4.	[7]	<ul style="list-style-type: none"> <li>• Classification</li> </ul>	<ul style="list-style-type: none"> <li>• NN</li> <li>• Decision Tree</li> <li>• SVM</li> <li>• Naïve Bayes</li> </ul>	<ul style="list-style-type: none"> <li>• Internal Assessment</li> <li>• Student Demographic</li> </ul>
5.	[8]	<ul style="list-style-type: none"> <li>• Classification</li> <li>• Regression</li> </ul>	<ul style="list-style-type: none"> <li>• Decision Tree</li> <li>• CHAID</li> <li>• Multiple Regression</li> </ul>	<ul style="list-style-type: none"> <li>• Internal Assessment</li> <li>• Social Network Interaction</li> <li>• Psychometric Factors</li> </ul>
6.	[9]	<ul style="list-style-type: none"> <li>• Classification</li> </ul>	<ul style="list-style-type: none"> <li>• C4.5</li> </ul>	<ul style="list-style-type: none"> <li>• CGPA</li> <li>• Student Demographic</li> <li>• Social Network Interaction</li> </ul>
7.	[10]	<ul style="list-style-type: none"> <li>• Classification</li> </ul>	<ul style="list-style-type: none"> <li>• C4.5</li> </ul>	<ul style="list-style-type: none"> <li>• Internal Assessment</li> <li>• Student Demographic</li> <li>• Social Network Interaction</li> </ul>

Berdasarkan dari studi literatur yang dilakukan, kami akan mengajukan untuk melakukan penelitian yang fokus dalam memprediksikan hasil ujian akhir nasional siswa SMAK Ora et Labora BSD karena belum dilakukan oleh penelitian sebelumnya.

### 3. Metodologi

Berdasarkan dari permasalahan yang ada, maka akan dilakukan penelitian dengan menggunakan data mining. Teknik yang digunakan adalah klasifikasi dan metodenya adalah Decision Tree C4.5. Ada dua data set yang akan digunakan, yaitu data internal assessment dari siswa SMAK Ora et Labora angkatan 20 yang saat ini telah menjadi alumni sebagai Data Training yang nantinya akan mejadi model klasifikasi. Kemudian data set yang kedua adalah data internal assesment siswa SMAK Ora et Labora angkatan 22 yang saat ini sedang duduk di kelas 11 sebagai data yang akan diprediksikan. Kedua data tersebut akan diambil dari database sekolah dan memiliki periode dari semester 1 hingga semester 3. Pengumpulan data yang dibutuhkan dilakukan dengan cara wawancara dan oberservasi langsung ke lokasi.

Dari data yang didapatkan, terdapat jumlah record sebesar 100 data murid untuk angkatan 20 yang akan berperan sebagai data training. Data tersebut terdiri dari 53 murid yang berasal dari jurusan IPS, dan 47 murid yang berasal dari jurusan MIPA. Untuk data testing, terdapat jumlah record sebesar 84 data murid angkatan 22. Data tersebut terdiri dari 42 murid yang berasal dari jurusan IPS, dan 42 murid yang berasal dari jurusan MIPA.

#### 3.1. *Pembangunan Data Mining*

Dalam pembangunan Data Mining, digunakan metodologi Knowledge Discovery in Database (KDD). Tahapan KDD adalah sebagai berikut:

##### 1. **Data Preprocessing**

Tahapan ini bertujuan untuk pembentukan data bersih dari data mentah yang didapatkan dari hasil observasi dan wawancara dengan pihak SMAK Ora et Labora. Data bersih diperlukan agar nantinya data yang digunakan untuk proses analisis sudah memenuhi kebutuhan dari peneliti. Tahapan Data Preprocessing ini terdiri dari proses Data Cleaning, Data Integration, Data Selection/Reduction, dan Data Transformation.

- **Data Cleaning**

Tahapan ini bertujuan untuk mencoba mengisi missing values, menghilangkan noise, mengidentifikasi pencilan, dan memperbaiki data yang tidak konsisten. Dari data yang sudah didapatkan, data internal assessment sudah disusun dengan cukup baik oleh pihak sekolah. Data nilai-nilai mata pelajaran sudah lengkap karena data tersebut harus masuk ke dalam nilai rapor per semester. Hanya saja terdapat data murid yang tidak lengkap pada kelas 10 dan kelas 11 yang dikarenakan siswa tersebut pindah sekolah atau tidak naik kelas. Selain itu ada juga kemungkinan pada saat kelas 11, sekolah mendapat murid pindahan sehingga data pada kelas 10 tidak ada atau tidak valid.

- **Data Integration**

Tahapan ini bertujuan untuk menyatukan data-data yang berasal dari sumber yang berbeda. Karena data yang digunakan berasal dari database yang sama, maka tahapan ini akan dilewatkan.

- **Data Selection/Reduction**

Tahapan ini bertujuan untuk melakukan pemilihan data yang berhubungan dengan analisis yang akan dilakukan. Data yang dipilih adalah data mata pelajaran yang memiliki hubungan secara langsung dengan mata pelajaran yang akan diujikan pada ujian akhir nasional. Pada murid jurusan MIPA, akan diambil mata pelajaran Matematika, Bahasa Indonesia, Bahasa Inggris, Fisika, Kimia, dan Biologi. Pada murid jurusan IPS, akan diambil mata pelajaran Matematika, Bahasa Indonesia, Bahasa Inggris, Ekonomi, Sosiologi, dan Geografi.

- **Data Transformation**

Tahapan ini bertujuan untuk mentransformasi data dan menggabungkannya menjadi suatu bentuk yang layak untuk dilakukan mining dengan membuat rangkuman atau operasi agregasi. Data yang sudah dipilih akan dirangkum menjadi suatu data set baru yang akan dieksekusi dalam proses data mining.

## 2. Data Mining

Tahapan ini adalah proses utama dan terpenting dimana metode pintar diaplikasikan untuk mengekstrak pola dari data. Tahapan ini menggunakan teknik classification dengan menggunakan metode Decision Tree/C4.5. Proses ini akan menggunakan software WEKA (Waikato Environment of Knowledge Analysis). Software WEKA yang digunakan adalah versi 3.8.2 yang merupakan versi terbaru. Sebenarnya ada versi yang lebih baru dari WEKA yaitu 3.9.2 namun versi ini sedang berada pada tahap development.

### Klasifikasi

Klasifikasi adalah suatu bentuk dari analisis data yang mengekstraksi model yang menjelaskan kelas data yang penting. Klasifikasi sendiri memiliki proses dua langkah, yaitu learning step (dimana model klasifikasi dibuat) dan classification step (dimana model digunakan untuk memprediksikan class label dari data yang ada). Learning step menghasilkan sejumlah record yang disebut data training yang terdiri dari kelas untuk record. Klasifikasi ini bertujuan untuk menghasilkan model dari data training yang membedakan record ke dalam beberapa kategori yang sesuai. Kemudian model tersebut akan digunakan untuk mengklasifikasikan record yang belum diketahui kelasnya. Pada data training yang ada, terdapat atribut-atribut berupa nilai mata pelajaran yang diambil dari semester 1 hingga semester 3. Dari atribut tersebut akan ditentukan kategori nilai akhir siswa pada akhir masa belajarnya. Kategori yang ditetapkan adalah AA(Above Average) dan BA (Below Average). Penentuan kategori ini adalah berdasarkan perbandingan nilai Ujian Akhir Nasional masing-masing siswa terhadap rata-rata dari seluruh nilai Ujian Akhir Nasional siswa angkatan 20.

## 3. Pattern Evaluation

Tahapan ini bertujuan untuk mengidentifikasi pola yang benar-benar menarik dalam merepresentasikan pengetahuan berdasarkan pada interestingness measures. Dalam tahap ini akan digunakan tools Confusion Matrix untuk mengevaluasi model klasifikasi untuk memperkirakan objek yang benar atau salah.

**Tabel 3.1** Format Confusion Matrix

Classification	Predicted Class	
	Class = Yes	Class = No
Class = Yes	(true positive-TP)	(false negative-FN)
Class = No	(false positive-FP)	(true negative-TN)

Evaluasi dan validasi akan diukur menggunakan rumus akurasi, precision, recall, dan F-measures. Pengukuran serta rumus-rumusnya dirangkum dalam tabel berikut ini.

**Tabel 3.2** Formula Evaluasi Klasifikasi

Measures	Formula
Akurasi	$Akurasi = \frac{TP + TN}{TP + TN + FP + FN}$
Precision	$Precision = \frac{TP}{TP + FP}$
Recall	$Recall = \frac{TP}{TP + FN}$
f-measures	$F = \frac{2 \times precision \times recall}{precision + recall}$

**4. Knowledge Presentation**

Tahapan ini bertujuan untuk membuat teknik visualisasi dan representasi pengetahuan digunakan untuk memberikan hasil mining kepada pihak SMAK Ora et Labora nantinya.

**4. Hasil dan Pembahasan**

Data training yang berasal dari data angkatan 20 murid SMAK Ora et Labora digunakan sebagai classification model. Kemudian model yang sudah dibuat akan dijadikan acuan untuk menguji data testing yang sudah disiapkan. Data testing yang digunakan berasal dari data angkatan 22 murid SMAK Ora et Labora. Data training dan data testing memiliki isi atribut yang sama. Namun perbedaannya atribut CAT (kategori) yang berada pada data testing akan dikosongkan sebagai yang nantinya akan diprediksi nilainya.

File data training dan data testing yang digunakan dalam software WEKA memiliki format .arff (Attribute-Relation File Format). Format data tersebut bisa didapatkan dengan mengkonversikan file excel menggunakan software WEKA.

**Pembahasan Hasil Data Mining Jurusan IPS**

Pertama akan dibahas dari jurusan IPS terlebih dahulu. Berikut ini adalah hasil dari proses data mining yang dilakukan.

```

Correctly Classified Instances      50          94.3396 %
Incorrectly Classified Instances    3           5.6604 %
Kappa statistic                    0.8868
Mean absolute error                 0.1
Root mean squared error             0.2236
Relative absolute error             20.005 %
Root relative squared error         44.7272 %
Total Number of Instances          53
    
```

**Gambar 4.1** Hasil Evaluasi Data Training Jurusan IPS

Dari proses data mining yang dijalankan, data training siswa jurusan IPS menghasilkan akurasi yang cukup baik yaitu sebesar 94.3396%. Kemudian didapatkan juga data sebagai berikut yang dirangkum dalam bentuk tabel.

**Tabel 4.1** Hasil Pengukuran Data Training Jurusan IPS

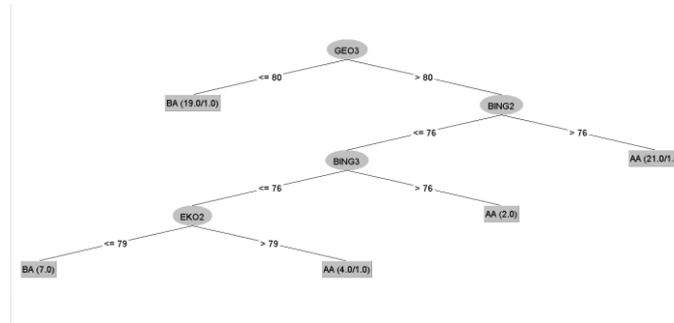
Class	TP Rate	FP Rate	Precision	Recall	F-Measure
AA	0.962	0.074	0.926	0.962	0.943
BA	0.926	0.038	0.962	0.926	0.943
Avg.	0.943	0.056	0.944	0.943	0.943

Berikut ini adalah Confusion Matrix yang dihasilkan dari proses data mining terhadap data training siswa jurusan IPS SMAK Ora et Labora.

**Tabel 4.2** Confusion Matrix Data Training Jurusan IPS

Classification	Predicted Class	
	Class = AA	Class = BA
Class = AA	25	1
Class = BA	2	25

Dihasilkan juga sebuah decision tree dari hasil data mining terhadap data training.



**Gambar 4.2** Decision Tree Data Training Jurusan IPS

Dari tree yang dihasilkan, bisa ditarik beberapa kesimpulan mengenai atribut atau mata pelajaran mana yang paling berpengaruh dalam hasil belajar siswa nantinya. Dari tree bisa dilihat bahwa mata pelajaran Geografi, Bahasa Inggris, dan Ekonomi adalah yang paling berpengaruh dalam hasil belajar siswa. Sehingga nantinya untuk ketiga mata pelajaran ini memerlukan perhatian lebih demi meningkatkan prestasi siswa.

### Pembahasan Hasil Data Mining Jurusan MIPA

Berikutnya akan dibahas hasil dari proses data mining yang dilakukan pada data training murid jurusan MIPA.

Correctly Classified Instances	45	95.7447 %
Incorrectly Classified Instances	2	4.2553 %
Kappa statistic	0.9117	
Mean absolute error	0.0783	
Root mean squared error	0.1979	
Relative absolute error	16.2408 %	
Root relative squared error	40.3299 %	
Total Number of Instances	47	

**Gambar 4.3** Hasil Evaluasi Data Training Jurusan IPA

Dari proses data mining yang dijalankan, data training siswa jurusan MIPA menghasilkan akurasi yang cukup baik yaitu sebesar 95.7447%. Kemudian didapatkan juga data sebagai berikut yang dirangkum dalam bentuk tabel.

**Tabel 4.3** Hasil Pengukuran Data Training Jurusan IPS

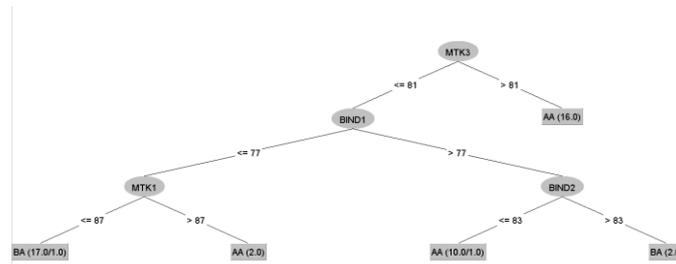
Class	TP Rate	FP Rate	Precision	Recall	F-Measure
AA	0.964	0.053	0.964	0.964	0.964
BA	0.947	0.036	0.947	0.947	0.947
Avg.	0.957	0.046	0.957	0.957	0.957

Berikut ini adalah Confusion Matrix yang dihasilkan dari proses data mining terhadap data training siswa jurusan MIPA SMAK Ora et Labora.

**Tabel 4.4** Confusion Matrix Data Training Jurusan IPA

Classification	Predicted Class	
	Class = AA	Class = BA
Class = AA	27	1
Class = BA	1	18

Dihasilkan juga sebuah decision tree dari hasil data mining terhadap data training.



**Gambar 4.4** Decision Tree Data Training Jurusan IPA

Dari tree yang dihasilkan, bisa ditarik beberapa kesimpulan mengenai atribut atau mata pelajaran mana yang paling berpengaruh dalam hasil belajar siswa nantinya. Dari tree bisa dilihat bahwa mata pelajaran Matematika dan Bahasa Indonesia adalah yang paling berpengaruh dalam hasil belajar siswa. Sehingga nantinya untuk ketiga mata pelajaran ini memerlukan perhatian lebih demi meningkatkan prestasi siswa.

**Pembahasan Hasil Prediksi Jurusan IPS**

Berdasarkan model yang didapatkan dari data training siswa jurusan IPS, dihasilkan sebuah file .arff yang berisikan hasil prediksi untuk data testing. Hasil tersebut kemudian dimasukkan ke dalam file data testing dengan cara mengisi kolom kategori yang kosong. Kemudian file tersebut kembali dikonversikan ke dalam file .arff agar bisa dijalankan di software WEKA.

Berikut ini adalah hasil pengujian akurasi dari hasil yang didapatkan.

```

Correctly Classified Instances      41          97.619 %
Incorrectly Classified Instances    1           2.381 %
Kappa statistic                    0.9522
Mean absolute error                 0.0452
Root mean squared error             0.1504
Relative absolute error             9.1266 %
Root relative squared error        30.2163 %
Total Number of Instances          42
    
```

**Gambar 4.5** Hasil Evaluasi DataTesting Jurusan IPS

Dari proses pengujian yang dijalankan, data testing siswa jurusan IPS menghasilkan akurasi yang cukup baik yaitu sebesar 97.619%. Kemudian didapatkan juga data sebagai berikut yang dirangkum dalam bentuk tabel.

**Tabel 4.5** Hasil Pengukuran Data Testing Jurusan IPS

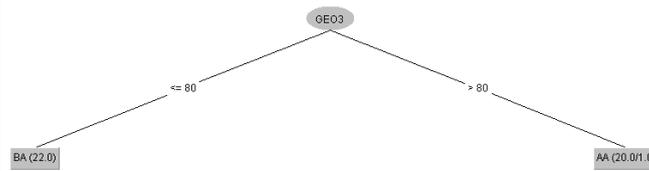
Class	TP Rate	FP Rate	Precision	Recall	F-Measure
AA	1.000	0.043	0.950	1.000	0.974
BA	0.957	0.000	1.000	0.957	0.978
Avg.	0.976	0.020	0.977	0.976	0.976

Berikut ini adalah Confusion Matrix yang dihasilkan dari proses pengujian terhadap data testing siswa jurusan IPS SMAK Ora et Labora.

**Tabel 4.6** Confusion Matrix Data Testing Jurusan IPS

Classification	Predicted Class	
	Class = AA	Class = BA
Class = AA	19	0
Class = BA	1	22

Dihasilkan juga sebuah decision tree dari hasil data mining terhadap data testing.



**Gambar 4.6** Decision Tree Data Testing Jurusan IPS

Dari tree yang dihasilkan, bisa ditarik beberapa kesimpulan mengenai atribut atau mata pelajaran mana yang paling berpengaruh dalam hasil belajar siswa nantinya. Dari tree bisa dilihat bahwa mata pelajaran Geografi adalah yang paling berpengaruh dalam hasil belajar siswa. Sehingga nantinya untuk mata pelajaran ini memerlukan perhatian lebih demi meningkatkan prestasi siswa.

**Pembahasan Hasil Prediksi Jurusan MIPA**

Berdasarkan model yang didapatkan dari data training siswa jurusan MIPA, dihasilkan sebuah file .arff yang berisikan hasil prediksi untuk data testing. Hasil tersebut kemudian dimasukkan ke dalam file data testing dengan cara mengisi kolom kategori yang kosong. Kemudian file tersebut kembali dikonversikan ke dalam file .arff agar bisa dijalankan di software WEKA.

Berikut ini adalah hasil pengujian akurasi dari hasil yang didapatkan.

```

Correctly Classified Instances      41          97.619 %
Incorrectly Classified Instances    1           2.381 %
Kappa statistic                    0.9517
Mean absolute error                 0.0451
Root mean squared error             0.1502
Relative absolute error             9.2018 %
Root relative squared error         30.3482 %
Total Number of Instances          42
  
```

**Gambar 4.7** Hasil Evaluasi Data Testing Jurusan IPA

Dari proses pengujian yang dijalankan, data testing siswa jurusan MIPA menghasilkan akurasi yang cukup baik yaitu sebesar 97.619%. Kemudian didapatkan juga data sebagai berikut yang dirangkum dalam bentuk tabel.

**Tabel 4.7** Hasil Pengukuran Data Testing Jurusan IPA

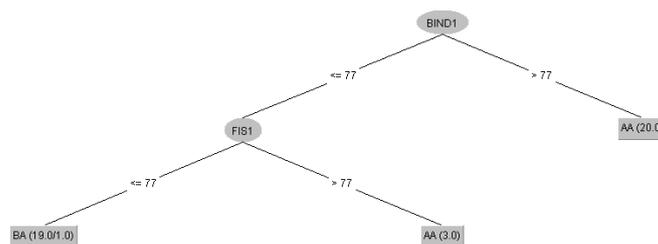
Class	TP Rate	FP Rate	Precision	Recall	F-Measure
BA	1.000	0.042	0.947	1.000	0.973
AA	0.958	0.000	1.000	0.958	0.979
Avg.	0.976	0.018	0.977	0.976	0.976

Berikut ini adalah Confusion Matrix yang dihasilkan dari proses pengujian terhadap data testing siswa jurusan IPS SMAK Ora et Labora.

**Table 4.8** Confusion Matrix Data Testing Jurusan IPA

Classification	Predicted Class	
	Class = BA	Class = AA
Class = BA	18	0
Class = AA	1	23

Dihasilkan juga sebuah decision tree dari hasil data mining terhadap data testing.



**Gambar 4.8** Decision Tree Data Testing Jurusan IPA

Dari tree yang dihasilkan, bisa ditarik beberapa kesimpulan mengenai atribut atau mata pelajaran mana yang paling berpengaruh dalam hasil belajar siswa nantinya. Dari tree bisa dilihat bahwa mata pelajaran Bahasa Indonesia dan Fisika adalah yang paling berpengaruh dalam hasil belajar siswa. Sehingga nantinya untuk kedua mata pelajaran ini memerlukan perhatian lebih demi meningkatkan prestasi siswa.

### 5. Kesimpulan dan Saran

Pada akhirnya didapatkan hasil prediksi yang bisa membantu pihak SMAK Ora et Labora untuk meningkatkan pembelajaran pada mata pelajaran yang dianggap membawa pengaruh dalam hasil belajar siswa. Pada jurusan IPS dihasilkan keputusan bahwa mata pelajaran Geografi perlu diperhatikan untuk mendukung prestasi siswa. Sedangkan pada jurusan MIPA dihasilkan keputusan bahwa mata pelajaran Bahasa Indonesia dan Fisika perlu diperhatikan untuk mendukung prestasi siswa. Dari keputusan yang ditarik ini, diharapkan pihak sekolah bisa melakukan tindakan pencegahan untuk mendongkrak hasil belajar siswa sesuai dengan jurusan masing-masing.

Penelitian ini hanya terbatas pada pembahasan hasil belajar siswa berdasarkan nilai mata pelajaran yang akan diujikan pada Ujian Akhir Nasional. Banyak atribut lainnya yang bisa digunakan untuk memprediksi hasil belajar siswa. Faktor seperti kegiatan ekstrakurikuler, dan faktor yang berhubungan dengan demografi siswa bisa digunakan untuk kedepannya. Sedangkan untuk metode yang digunakan, bisa digunakan algoritma lain yang berbeda.

#### Ucapan Terima Kasih

Segala pujian, kehormatan, dan kemuliaan bagi Tuhan Yesus Kristus atas rahmat dan berkat-Nya karena senantiasa menguatkan peneliti untuk menyelesaikan jurnal ini. Dalam menyusun jurnal ini, banyak orang telah memberikan motivasi, saran, dan dukungan bagi peneliti. Dalam kesempatan yang berharga ini, peneliti bermaksud menyampaikan rasa terima kasih dan penghargaannya kepada mereka semua.

Pertama, penghargaan terdalam dari peneliti diberikan kepada anggota keluarga tercinta atas cinta, doa, dan dukungan yang tak berkesudahan. Juga jurnal ini tidak akan mungkin terjadi tanpa bantuan, dukungan dan kesabaran penasihat saya Dr. Sani M. Isa, S.Si., M.Kom. atas pengawasan, saran, dan bimbingannya sejak tahap awal penelitian ini serta memberi saya pengalaman luar biasa selama beberapa bulan terakhir. Dia juga membantu saya dengan sabar menyelesaikan jurnal ini dengan memberikan saran, bimbingan, dan koreksi hingga jurnal ini selesai. Penghargaan terbesar peneliti juga diberikan kepada Dr. Eng. Antoni Wibowo, S.Si., M.Kom., M.Eng dan Abba Suganda Girsang ST, M.Cs., Ph.D sebagai penguji atas saran, pengawasan, dan kontribusi penting mereka dalam peningkatan hasil jurnal ini. Saya berterima kasih kepada kepala sekolah SMAK Ora et Labora, Ibu Maria Goretti yang mengizinkan saya melakukan penelitian di sana.

Akhirnya, saya ingin mengucapkan terima kasih kepada semua orang yang penting bagi keberhasilan realisasi jurnal ini. Jurnal ini jauh dari sempurna, tetapi diharapkan bermanfaat tidak hanya bagi peneliti, tetapi juga bagi pembaca. Untuk alasan ini, penulis akan menerima segala saran dan kritik yang konstruktif.

## References

- [1] S. K. Mohamad and Z. Tasir, "Educational Data Mining: A Review," *Procedia - Soc. Behav. Sci.*, vol. 97, pp. 320–324, 2013.
- [2] X. Wu et al., *Top 10 algorithms in data mining*, vol. 14, no. 1. 2008.
- [3] A. M. Shahiri, W. Husain, and N. A. Rashid, "A Review on Predicting Student's Performance Using Data Mining Techniques," *Procedia Comput. Sci.*, vol. 72, pp. 414–422, 2015.
- [4] R. Asif, A. Merceron, S. A. Ali, and N. G. Haider, "Analyzing undergraduate students' performance using educational data mining," *Comput. Educ.*, vol. 113, pp. 177–194, 2017.
- [5] C. Burgos, M. L. Campanario, D. de la Peña, J. A. Lara, D. Lizcano, and M. A. Martínez, "Data mining for modeling students' performance: A tutoring action plan to prevent academic dropout," *Comput. Electr. Eng.*, vol. 66, pp. 541–556, 2018.
- [6] R. Campagni, D. Merlini, R. Sprugnoli, and M. C. Verri, "Data mining models for student careers," *Expert Syst. Appl.*, vol. 42, no. 13, pp. 5508–5521, 2015.
- [7] E. B. Costa, B. Fonseca, M. A. Santana, F. F. de Araújo, and J. Rego, "Evaluating the effectiveness of educational data mining techniques for early prediction of students' academic failure in introductory programming courses," *Comput. Human Behav.*, vol. 73, pp. 247–256, 2017.
- [8] H. Susanto and Sudiyanto, "Data Mining Untuk Memprediksi Prestasi Siswa Berdasarkan Sosial Ekonomi, Motivasi, Kedisiplinan Dan Prestasi Masa Lalu," *J. Pendidik. Vokasi*, vol. 4, no. 2, pp. 222–231, 2014.
- [9] D. Kurniawan, W. Wibowo, and P. Y. Astuti, "Pemanfaatan Educational Data Mining (Edm) Untuk Memprediksi Masa Studi Mahasiswa Menggunakan Algoritma C4.5 (Studi Kasus: Ti-S1 Udinus)," *Momentum*, vol. 12, pp. 48–52, 2007.
- [10] D. Maurina and A. Zainul Fanani, "Penerapan Data Mining Untuk Rekomendasi Beasiswa Pada Sma Muhammadiyah Gubug Menggunakan," pp. 1–5, 2016.