

KLASIFIKASI KELULUSAN UJI KOMPETENSI KEAHLIAN PADA SMK MALAKA JAKARTA TIMUR

Irlon.

Program Studi Teknik Informatika, FTI, Institut Teknologi Budi Utomo Jakarta,
dahil.irlon@gmail.com

Abstrak

SMK Malaka Jakarta Timur, sebuah lembaga pendidikan kejuruan, menyelenggarakan Uji Kompetensi Keahlian setiap tahun. Meramalkan tingkat kelulusan menjadi langkah penting dalam menangani masalah kelulusan Uji Kompetensi Keahlian di SMK Malaka Jakarta Timur, di mana lebih dari 5% siswa tidak lulus setiap tahun. Penelitian ini memanfaatkan basis data siswa SMK Malaka Jakarta Timur dari tahun 2021 hingga 2023. Sepuluh variabel atribut dipilih, termasuk jenis kelamin, tempat lahir, pekerjaan orang tua, asal sekolah, kepemilikan laptop, jarak rumah, serta nilai Ujian Nasional (Matematika, IPA, Bahasa Inggris, Bahasa Indonesia), dan satu label atribut hasil uji kompetensi keahlian. Data dianalisis menggunakan Algoritma C4.5, yang menghasilkan akurasi sebesar 88,83%. Dengan menerapkan Algoritma C4.5 berbasis forward selection, akurasi meningkat menjadi 92,26%.

Kata kunci : Algoritma C4.5 dan Forward Selection, Prediksi kelulusan, Uji Kompetensi Keahlian

1. PENDAHULUAN

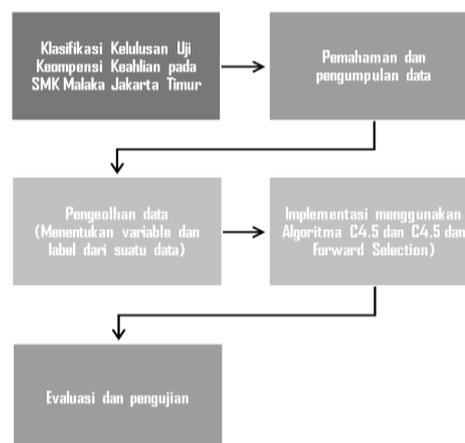
Sekolah Menengah Kejuruan (SMK) Malaka Jakarta Timur, sebagai lembaga pendidikan yang berfokus pada pengembangan keterampilan kejuruan, memiliki komitmen kuat dalam memastikan kelulusan siswa melalui Uji Kompetensi Keahlian (UKK). Ujian ini merupakan tolok ukur penting dalam menilai kemampuan siswa, baik dari segi teori maupun praktik, sesuai dengan bidang keahlian yang dipelajari. Mengingat pentingnya UKK dalam menentukan kelulusan dan kesiapan siswa memasuki dunia kerja, prediksi kelulusan menjadi langkah krusial yang harus dilakukan oleh pihak sekolah.

Menurut Peraturan Menteri Pendidikan dan Kebudayaan Nomor 3 Tahun 2017 tentang penilaian hasil belajar satuan pendidikan, penilaian hasil belajar siswa pada jenjang pendidikan menengah dilakukan melalui berbagai mekanisme, salah satunya adalah Uji Kompetensi Keahlian. Pedoman penyelenggaraan UKK tahun pelajaran 2016/2017 juga menggarisbawahi pentingnya ujian ini sebagai bagian integral dari proses penilaian hasil belajar siswa. Oleh karena itu, memastikan tingginya tingkat kelulusan siswa dalam UKK merupakan prioritas utama bagi SMK Malaka Jakarta Timur.

Namun, dalam beberapa tahun terakhir, lebih dari 5% siswa di SMK Malaka Jakarta Timur tidak berhasil lulus dalam UKK setiap tahunnya. Hal ini menimbulkan kekhawatiran terhadap efektivitas metode pengajaran dan persiapan ujian yang diterapkan. Oleh karena itu, diperlukan suatu pendekatan yang lebih terstruktur dan ilmiah untuk memahami faktor-faktor yang mempengaruhi kelulusan siswa dan untuk mengembangkan model prediktif yang dapat membantu meningkatkan tingkat kelulusan..

2. METODOLOGI

2.1 Kerangka Pemikiran



Gambar 1 : Kerangka Pemikiran
Sumber : Olahan data mandiri

Tabel 1 Pemilihan Atribut

No	Attribute	Content	Note
1	JK	Laki-laki, Perempuan	Jenis Kelamin
2	Tempat	SMR, LSMR	Tempat tanggal lahir SMR = Samarinda LSMR= Luar Samarinda
3	Pekerjaan	PNS, TB, KS, WR,	Pekerjaan Orang Tua PNS = PNS/TNI/Polri TB = Tidak Bekerja KS = Karyawan Swasta WR = Wiraswasta
4	Asal	NK, NL, SK, SL	Asal Sekolah NK= Negeri dalam Kota NL = Negeri Luar Kota SK = Swasta dalam Kota SL = Swasta Luar Kota
5	Laptop	Yes, No	Kepemilikan laptop Yes= punya laptop No= tidak punya laptop
6	Jarak	1-	Jarak dari rumah ke sekolah dalam satuan KM
7	MTK	1-10	Nilai UN SMP
8	IPA	1-10	Nilai UN SMP
9	Inggris	1-10	Nilai UN SMP
10	Indonesia	1-10	Nilai UN SMP
11	Hasil	Lulus, Tidak Lulus	Hasil Akhir

Sumber: Olehan Data Mandiri

2.2 Langkah – langkah Penelitian

2.2.1 Bisnis Understanding Phase

Pada tahap business understanding phase adalah tentang pemahaman mengenai tujuan dari kegiatan data mining yang akan dilakukan serta kebutuhan dari perspektif bisnis. SMK Malaka Jakarta Timur sebagai lembaga pendidikan kejuruan yang diatur dalam peraturan menteri pendidikan dan kebudayaan Nomor 3 Tahun 2017 tentang penilaian hasil belajar satuan pendidikan, dan pedoman penyelenggaraan UKK (Uji Kompetensi Keahlian) tahun pelajaran 2021/2023 peserta didik Sekolah Menengah Kejuruan harus menyelesaikan Uji Kompetensi Keahlian yang akan menjadi indikator ketercapaian standar kompetensi lulusan.

2.2.2 Data Understanding Phase

Pada tahapan ini akan dibahas mengenai data yang akan digunakan untuk proses data mining. Sesuai dengan tahapan sebelumnya data mining akan dilakukan berdasarkan data Uji Kompetensi Keahlian (UKK) peserta didik di SMK Malaka Jakarta timur.

2.3.2 Data Prepartion

Pada tahapan data preparation ini akan mencakup semua kegiatan untuk mempersiapkan data yang akan dimasukkan ke dalam alat pemodelan atau pada fase selanjutnya, dimana data tersebut merupakan pengolahan dari data mentah awal. Setelah data didapatkan lalu dilakukan clean data atau pembersihan data yang kurang, kosong atau noise, dengan cara memperbaiki, atau menghapus data sehingga data dapat diproses. Dalam data preparation phase ini dilakukan proses penentuan variable dan cleaning data seperti di bawah ini :

Proses cleaning data pada tahap ini adalah pembersihan data yang bersifat manual. Dimana proses pembersihan data dilakukan di luar aplikasi. Data dibersihkan dari data yang memiliki missing value. Maksud dari missing value di sini contohnya adalah data yang tidak lengkap. Maksud dari data tidak lengkap contohnya ada field yang kosong dan apabila ada field yang kosong maka data tersebut akan dihilangkan atau ditambah secara manual pada data set yang disimpan dalam format CSV (Comma Separated Values) dengan menggunakan Microsoft office excel.

Tabel 2. Hasil *Cleaning*

JK	Tempat	Pekerjaan	Asal	Laptop	Jarak	MTK	IPA	Inggris	Indonesia	Hasil
Laki-Laki	SMR	PNS	NK	Yes	8	7.38	7	7.8	6.64	Lulus
Laki-Laki	SMR	TB	NL	No	2	7.38	7.18	7.7	6.64	Lulus
Laki-Laki	SMR	KS	NK	Yes	1	7	7.18	7.76	6.54	Lulus
Laki-Laki	LSMR	WR	NL	Yes	5	7.14	6.54	7.8	6.64	Lulus
Laki-Laki	SMR	KS	SK	Yes	3	7.14	6.54	7.68	6.64	Tidak Lulus
Laki-Laki	LSMR	KS	NK	Yes	4	6.78	6.54	7.48	6.64	Lulus
Laki-Laki	SMR	WR	NK	No	4	6.78	7.18	7.86	6.64	Tidak Lulus
Laki-Laki	SMR	WR	NK	Yes	1	7.38	6.54	7.58	6.64	Lulus
Laki-Laki	LSMR	PNS	NL	Yes	1	7.38	6.54	7.76	6.72	Lulus
Laki-Laki	SMR	KS	NK	Yes	2	7.64	6.54	7.86	6.82	Lulus
Laki-Laki	SMR	TB	NK	Yes	3	7	6.98	7.76	6.64	Lulus
Laki-Laki	SMR	WR	NK	No	2	6.78	6.98	7.38	6.82	Lulus
Laki-Laki	SMR	WR	NK	Yes	3	7.38	7.18	7.58	7	Tidak Lulus
Laki-Laki	SMR	WR	SK	No	4	7.64	7.18	7.58	6.72	Lulus
Laki-Laki	LSMR	WR	NK	Yes	3	6.78	6.82	7.86	6.64	Lulus
Laki-Laki	SMR	WR	NL	Yes	8	7	6.54	6.54	6.64	Lulus
Laki-Laki	SMR	WR	NK	No	10	6.78	6.54	7.7	6.64	Lulus
Laki-Laki	SMR	KS	NK	Yes	3	7	7.28	7.76	6.64	Lulus
Laki-Laki	SMR	KS	NK	Yes	10	6.78	6.54	6.54	6.64	Lulus

Sumber: Olehan Data Mandiri

2.3.3 Modeling

Permodelan masing-masing algoritma dilakukan dengan menggunakan 111 data, dimana dengan menggunakan algoritma yang telah ditentukan sebelumnya. Hal ini dilakukan dengan mengacu kepada bentuk data dan tujuan/hasil yang ingin dicapai pada penerapan data mining, yaitu untuk memprediksi kelulusan hasil uji kompetensi keahlian TKJ

2.3.4 Evaluation

Penelitian yang akan dilakukan dalam eksperimen ini adalah dengan menggunakan komputer untuk melakukan proses perhitungan terhadap model yang diusulkan. Proses eksperimen dan pengujian model menggunakan bagian dari dataset yang ada. Semua dataset kemudian diuji dengan metode yang diusulkan yaitu algoritma C4.5 dan forward selection untuk pemilihan atribut pada aplikasi Rapid miner 7 Open source berlisensi AGPL (GNU Affero General Public License) versi 3. Dari 349 data yang diuji menggunakan Confusion Matrix untuk melihat akurasi dari perhitungan menggunakan algoritma C4.5 terhadap data yang digunakan

2.3.5 Deployment

Dalam konteks pemodelan, untuk memperoleh hasil terbaik dari model, proses pemodelan dapat diulang dengan melakukan pelatihan ulang menggunakan penambahan data dan melakukan pemilihan fitur dengan forward selection untuk menentukan variabel yang tepat. Penambahan data dan pemilihan fitur merupakan langkah-langkah penting dalam meningkatkan kinerja model serta memastikan bahwa variabel yang digunakan dalam pemodelan adalah yang paling relevan dan memberikan kontribusi yang signifikan terhadap hasil prediksi.

3. ANALISA DAN PEMBAHASAN

3.1 Implementasi Data Mining

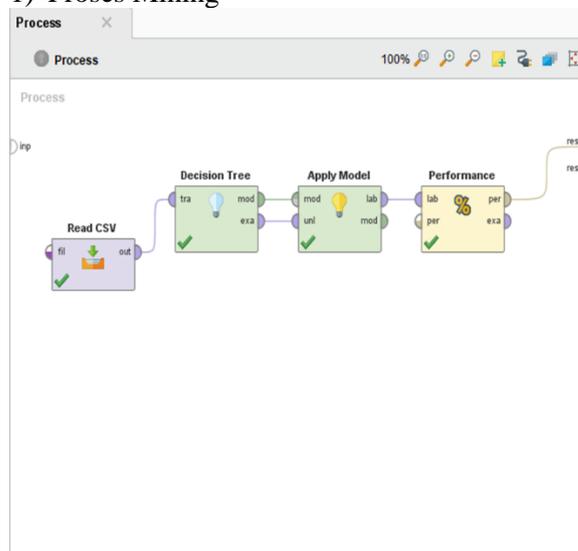
Pada penelitian ini menggunakan database SMK Malaka Jakarta Timur dari penerimaan peserta didik sejak 2021 - 2023 sebanyak 349 data record . Penelitian ini menggunakan analisis dengan algoritma C4.5 dan seleksi atribut dengan menggunakan Forward Selection dalam proses data mining yang dilakukan, untuk mengetahui atribut mana saja yang mempengaruhi hasil dari

evaluation performance, yang nantinya dapat dijelaskan pada sub bab di bawah ini.

3.1.1 Implementasi Algoritma C4.5

Dalam proses mining ini menggunakan 4 operator, yaitu Read CSV, Decision Tree, Apply Model, Performance

1) Proses Mining



Gambar 2. Proses Mining dengan Algoritma C4.5

Sumber ; Olahan Data Mandiri

2) Confusion Matrix

accuracy: 88.83%

	true Lulus	true Tidak Lulus	class precision
pred. Lulus	221	38	85.33%
pred. Tidak Lulus	1	89	98.89%
class recall	99.55%	70.00%	

Gambar 3. Hasil Accuracy dengan Algoritma C4.5
Sumber ; Olahan Data Mandiri

$$Accuracy = \frac{TP + TN}{P + N} = \frac{89 + 221}{(89 + 38 + 1 + 221)} = 0.8883 \times 100\% = 88.83\%$$

precision: 98.89% (positive class: Tidak Lulus)

	true Lulus	true Tidak Lulus	class precision
pred. Lulus	221	38	85.33%
pred. Tidak Lulus	1	89	98.89%
class recall	99.55%	70.00%	

Gambar 4. Hasil precision Algoritma C4.5
Sumber ; Olahan Data Mandiri

$$Precision = \frac{TP}{TP+FP} = \frac{89}{89+1} = 0.9889 \times 100\% = 98.89\%$$

recall: 70.08% (positive class: Tidak Lulus)			
	true Lulus	true Tidak Lulus	class precision
pred. Lulus	221	38	85.33%
pred. Tidak Lulus	1	89	98.89%
class recall	99.55%	70.08%	

Gambar 5. Hasil Recall Algoritma C4.5
Sumber ; Olahan Data Mandiri

$$Recall = \frac{TP}{TP+FN} = \frac{89}{89+38} = 0.7008 \times 100\% = 70.08\%$$

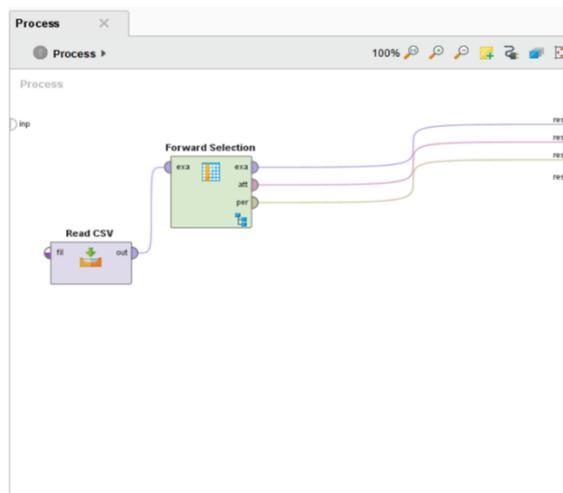


Gambar 6. Curva AUC Algoritma C4.5
Sumber ; Olahan Data Mandiri

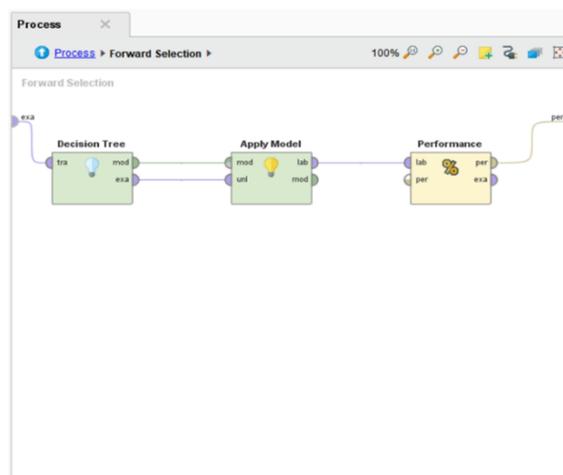
Kurva ROC (*Receiver Operating Characteristic*) adalah representasi grafis dari hubungan antara sensitivitas dan 1-spesifisitas. Nilai 0.889 pada kurva ROC menunjukkan tingkat akurasi model dalam membedakan antara dua kelas pada berbagai threshold. Semakin dekat nilai AUC (*Area Under the Curve*) dengan 1, semakin baik kemampuan model dalam membedakan antara kelas positif dan negatif. Dengan demikian, nilai 0.889 menunjukkan bahwa model memiliki tingkat akurasi yang baik dalam klasifikasi

3.1.2 Implementasi Algoritma C4.5 Forward Selection

1) Proses Mining



Gambar 7. Proses Mining Forward Selection
Sumber ; Olahan Data Mandiri



Gambar 8. Proses Mining Algoritma C4.5 dengan Forward Selection
Sumber ; Olahan Data Mandiri

2) Confusion Matrix

accuracy: 92.26%			
	true Lulus	true Tidak Lulus	class precision
pred. Lulus	207	12	94.52%
pred. Tidak Lulus	15	115	88.46%
class recall	93.24%	90.55%	

Gambar 9. Hasil Accuracy Algoritma C4.5 dengan Forward Selection
Sumber ; Olahan Data Mandiri

$$Accuracy = \frac{TP+TN}{P+N} = \frac{115+207}{(115+12+15+207)}$$

precision: 88.46% (positive class: Tidak Lulus)			
	true Lulus	true Tidak Lulus	class precision
pred. Lulus	207	12	94.52%
pred. Tidak Lulus	15	115	88.46%
class recall	93.24%	90.55%	

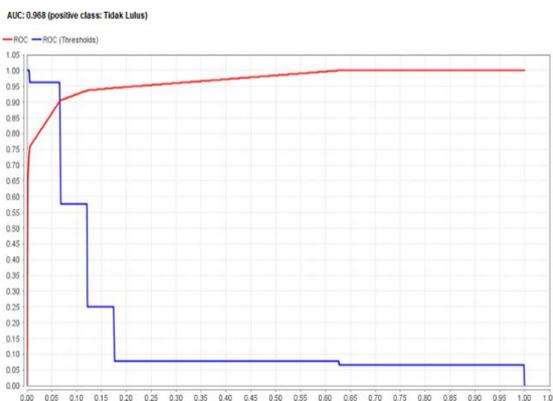
Gambar 10. Hasil precision Algoritma C4.5 dengan Forward Selection
Sumber ; Olahan Data Mandiri

$$Precision = \frac{TP}{TP+FP} = \frac{115}{115+15} = 0.8846 \times 100\% = 88.46\%$$

recall: 90.55% (positive class: Tidak Lulus)			
	true Lulus	true Tidak Lulus	class precision
pred. Lulus	207	12	94.52%
pred. Tidak Lulus	15	115	88.46%
class recall	93.24%	90.55%	

Gambar 11. Hasil Recall Algoritma C4.5 dengan Forward Selection
Sumber ; Olahan Data Mandiri

$$Recall = \frac{TP}{TP+FN} = \frac{115}{115+12} = 0.9055 \times 100\% = 90.55\%$$



Gambar 11. Curva AUC Algoritma C4.5 dengan Forward Selection
Sumber ; Olahan Data Mandiri

Kurva ROC (Receiver Operating Characteristic) adalah representasi grafis dari hubungan antara sensitivitas dan 1-spesifisitas. Nilai 0.889 pada kurva ROC menunjukkan tingkat akurasi model dalam membedakan antara dua kelas pada berbagai threshold. Semakin dekat nilai AUC (Area Under the Curve) dengan 1, semakin baik kemampuan model dalam membedakan antara kelas positif dan negatif. Dengan demikian, nilai 0.968 menunjukkan bahwa model memiliki tingkat akurasi yang baik dalam klasifikasi

3) Forward Selection

Dalam hasil uji coba algoritma C4.5 berbasis forward selection menggunakan data set 349 record memberikan 5 atribut terbaik yang bernilai 1, dan 5 atribut bernilai 0. Atribut hasil seleksi fitur adalah Pekerjaan, Laptop, Jarak, MTK, Inggris dengan tingkat akurasi 92.26%.

Table 3. Table Attribute weight (Forward Selection)

Attribute	weight
JK	0
Tempat	0
Pekerjaan	1
Asal	0
Laptop	1
Jarak	1
MTK	1
IPA	0
Inggris	1
Indonesia	0

Sumber ; Olahan Data Mandiri

Bobot yang diberikan kepada atribut-atribut dalam tabel ini menunjukkan pentingnya masing-masing atribut dalam model yang sedang dikembangkan. Atribut dengan bobot 1 dianggap signifikan dan berkontribusi terhadap model, sedangkan atribut dengan bobot 0 dianggap tidak signifikan dan tidak berkontribusi terhadap model.

4. KESIMPULAN

Setelah melakukan penelitian dan pengujian untuk memprediksi kelulusan Uji Kompetensi Keahlian pada program keahlian TKJ di SMK Malaka Jakarta Timur dengan menggunakan Algoritma C4.5, maka didapatkan kesimpulan, yaitu :

- 1) Setelah melakukan proses data mining terhadap data peserta didik pada SMK Malaka Jakarta Timur menggunakan algoritma C4.5 menghasilkan nilai accuracy 88.83%, dan saat menggunakan algoritma dengan forward selection menghasilkan nilai accuracy 92.26% yang artinya pemilihan atribut dengan forward selection meningkatkan nilai accuracy.
- 2) Algoritma C4.5 dengan forward selection menghasilkan nilai kurva ROC 0.968 yang artinya Excellent Classification.
- 3) Dengan bantuan forward selection pada 349 data set dihasilkan 5 atribut yang berpengaruh untuk memprediksi

kelulusan Uji Kompetensi Keahlian yaitu Pekerjaan, Laptop, Jarak, MTK, dan Inggris.

- 4) Dengan adanya penelitian ini dapat membantu pihak SMK Malaka Jakarta Timur untuk memberikan peringatan dini bagi peserta didik yang kemungkinan tidak lulus Uji Kompetensi Keahlian (UKK) dan sebagaimana yang diharapkan. program keahlian dalam membuat kebijakan untuk bisa meningkatkan kelulusan.

DAFTAR PUSTAKA

- Baradwaj, B. K., & Pal, S. (2012). Mining Educational Data to Analyze Students Performance. *International Journal of Advanced Computer Science and Applications*, 63-69.
- Chen, F., Deng, P., Wan, J., Zhang, D., Vasilakos, A. V., & Rong, X. (2015). Data Mining for the Internet of Things: Literature Review and Challenges. *International Journal of Distributed Sensor Networks*, 2.
- Dhika, H., & Destiwati, F. (2015). Application of data mining algorithm to recipient of motorcycle installment. *Journal ComTech*, 569-579.
- Direktur Pembinaan SMK. (2017). Pedoman Penyelenggaraan UKK Tahun Pelajaran 2016/2017. Jakarta: Depdiknas.
- Dongming, L., Yan, L., Chao, Y., Chaoran, L., Huan, L., & Lijuan, Z. (2016). The Application of Decision Tree C4.5 Algorithm to Soil Quality Grade Forecasting Model. *First IEEE International Conference on Computer Communication and the Internet*, 552-555.