

Optimasi Algoritma C4.5 Menggunakan Metode Adaboost Classification Pada Klasifikasi Nilai Mahasiswa Studi Kasus: Universitas Muhammadiyah Kalimantan Timur

Optimizing the C4.5 Algorithm Using Adaboost Classification Method for Student Grade Classification Case Study: Muhammadiyah University of East Kalimantan

Suci Mawaddah
Wawan Joko Pranoto
Faldi

Program Studi SI Teknik Informatika
Universitas Muhammadiyah Kalimantan Timur
Jalan Ir. Juanda No.15 Samarinda Ulu, Kalimantan Timur, Indonesia
1911102441089@umkt.ac.id

ABSTRAK

Penelitian ini membahas tentang klasifikasi nilai mahasiswa dengan menggunakan optimasi algoritma C4.5 menggunakan Adaboost Classification. Dengan adanya permasalahan yang dihadapi yaitu, penurunan nilai mahasiswa yang drastis, maka tujuan penelitian ini untuk mengetahui indikator yang mempengaruhi penurunan nilai mahasiswa dan meningkatkan persentase akurasi pada algoritma C4.5 menggunakan metode *Adaboost Classification*. Hasil pengujian awal dengan algoritma C4.5 menunjukkan akurasi sebesar 81% dalam klasifikasi nilai mahasiswa. Namun, akurasi tersebut perlu ditingkatkan. Oleh karena itu, penelitian ini menerapkan metode seleksi fitur dengan menambahkan metode *Adaboost Classification* untuk mengoptimalkan akurasi algoritma C4.5. Hasil pengujian menunjukkan bahwa dengan metode *Adaboost Classification*, akurasi dapat meningkat menjadi 85% dengan indikator yang berpengaruh antara lain progress, course completed, tugas 1, tugas 2 dan simbol sebagai kelas targetnya. Penelitian ini memberikan kontribusi dalam meningkatkan akurasi dengan mengoptimalkan algoritma C4.5 melalui metode Adaboost Classification serta dapat digunakan untuk meningkatkan system evaluasi nilai mahasiswa untuk meningkatkan kualitas pendidikan.

Kata kunci: Akurasi, Algoritma C4.5, Metode Adaboost Classification, Nilai Mahasiswa, Klasifikasi.

ABSTRACT

This study focuses on the classification of student grades using the optimization of the C4.5 algorithm using Adaboost Classification. The main objective is to identify the indicators influencing the drastic decrease in student grades and improve the accuracy percentage of the C4.5 algorithm using the Adaboost Classification method. The initial testing of the C4.5 algorithm showed an accuracy of 81% in classifying student grades. However, this accuracy needs to be improved. Therefore, this study applies feature selection methods by incorporating the Adaboost Classification method to optimize the accuracy of the C4.5 algorithm. The results of the testing indicate that with the Adaboost Classification method, the accuracy can be increased to 85%, with indicators such as progress, course completion, assignment 1, assignment 2, and symbols as the target class. This research contributes to enhancing the accuracy of student grade classification by optimizing the C4.5 algorithm through the Adaboost Classification method. The findings can be used to improve the student grade evaluation system at Muhammadiyah University of East Kalimantan and enhance the quality of education.

Keywords: Accuracy, C4.5 Algorithm, Adaboost Classification Method, Student Grades, Classification.

Pendahuluan

Pada perguruan tinggi, mahasiswa merupakan parameter penting dalam pergerakan bisnis proses dan evaluasi penyelenggaraan program studi (Atma &

Setyanto, 2018). Untuk menjaga hal yang berkaitan dengan prestasi, kompetensi, dan presensi mahasiswa maka institusi pendidikan khususnya Universitas Muhammadiyah Kalimantan Timur (UMKT) perlu

memberikan perhatian yang serius dalam evaluasi kinerja mahasiswa.

Dengan adanya data dari Mata Kuliah Dasar Umum (MKDU) UMKT salah satunya nilai mata kuliah Bahasa Indonesia, yang menunjukkan bahwa nilai mahasiswa UMKT dari angkatan 2020 ke angkatan 2021 mengalami penurunan drastic, maka untuk menjaga kualitas kinerja mahasiswa perlu adanya proses monitoring secara berkala sebagai proses tindak lanjutnya, sehingga menciptakan pengaruh besar untuk kinerja mahasiswa itu sendiri dalam akademik, serta berpengaruh juga pada rasio dan poin akreditasi perguruan tinggi tersebut.

Dalam rangka menggali informasi tersebut salah satu teknik yang dapat digunakan yakni, teknik data mining. Salah satu algoritma yang dapat digunakan, yaitu algoritma C4.5. Algoritma C4.5 adalah salah satu metode untuk membuat decision tree berdasarkan training data yang telah disediakan. Sebagai salah satu pengembangan dari ID3 yang diciptakan oleh Ross Quinlan, beberapa pengembangan yang dilakukan pada C4.5 antara lain bisa mengatasi missing value, bisa mengatasi data kontinyu, dan pruning (Wanto et al., 2020).

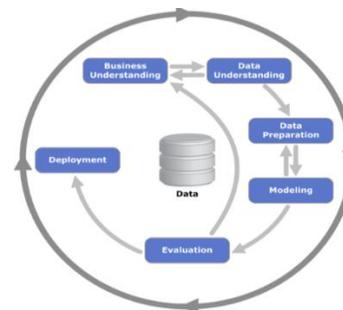
Penelitian ini dilakukan untuk mengatasi kerendahan akurasi pada algoritma C4.5 menggunakan metode Adaboost Classification dengan mengklasifikasikan data atribut yang diperoleh dari Mata Kuliah Dasar Umum (MKDU) Universitas Muhammadiyah Kalimantan Timur.

Berdasarkan penelitian sebelumnya beberapa penggunaan algoritma C4.5 pada kinerja mahasiswa dalam klasifikasi nilai, diantaranya : Penelitian yang dilakukan oleh (Rustam, 2020) pada kasus menentukan kelayakan mata kuliah konsentrasi didapatkan hasil akurasi terbaik 89.86% dalam pengklasifikasian kompetensi. Penelitian selanjutnya dengan kasus prediksi kelulusan mahasiswa dilakukan (Anggita & Ikamah, 2021) dengan mengoptimasi bobot atribut yang digunakan sebagai faktor kelulusan mahasiswa pada Algoritma C4.5 menggunakan optimasi PSO menghasilkan akurasi 90.73%, setelah dilakukan optimasi menggunakan algoritma PSO untuk membobotkan atribut pada algoritma C4.5 diperoleh akurasi sebesar 97.78%. Meskipun demikian, masih terdapat kerendahan nilai akurasi saat ditambahkan fitur optimasi pada algoritma C4.5. Penelitian yang dilakukan (Atma & Setyanto, 2018) dalam mengidentifikasi mahasiswa berpotensi drop out dengan membandingkan algoritma C4.5 dan KNN dengan penambahan fitur Forward Selection, performa KNN + Forward Selection lebih unggul dengan akurasi 98.34%, sedangkan algoritma C4.5 + Forward Selection memiliki akurasi 96.66%. Hal tersebut menjadi sebuah alasan untuk mengatasi kerendahan akurasi pada

algoritma C4.5 dengan melakukan penambahan metode seleksi atau fitur optimasi yang berbeda pada penelitian sebelumnya dengan menggunakan metode seleksi Adaboost Classification.

Metode

Adapun metode penelitian dari penelitian ini menggunakan Metodologi Cross Industry Standard Process for Data Mining (CRISP-DM). CRISP-DM memiliki 6 tahapan, yaitu seperti pada Gambar 1 (Komang Aditya Pratama et al., 2020):



Gambar 1 Metode Penelitian

1. Business Understanding

Business Understanding merupakan tahapan dalam mengenal proses bisnis berupa mengenal tujuan bisnis, kondisi dan permasalahan lapangan serta kebutuhannya.

2. Data Understanding

Data Understanding merupakan tahapan pengumpulan data, pemahaman terhadap data yang akan digunakan, data yang telah terkumpul dideskripsikan untuk diketahui mana yang menjadi atribut, kelas dan jenis tipe datanya; evaluasi kualitas data dan mengidentifikasi masalah yang berkaitan dengan kualitas data.

3. Data Preparation

Data Preparation merupakan tahapan pengolahan data mentah sehingga menjadi data yang lebih berarti. Penerapan data Preparation pada dataset dilakukan dalam beberapa tahap, antara lain:

- 1) Data Cleaning, yaitu tahap untuk membersihkan atribut-atribut dan data yang kosong.
- 2) Data Integration, yaitu menyatukan beberapa data yang berbeda menjadi satu data.
- 3) Data Reduction, yaitu mengurangi jumlah atribut dan data yang dianggap sudah terlalu banyak, juga melakukan penghapusan duplikat.
- 4) Data Conversion, yaitu mengklasifikasikan data menjadi beberapa kategori berdasarkan atributnya.

4. Modelling

Modelling merupakan tahapan penerapan algoritma untuk mencari, mengidentifikasi, dan menemukan pola.

Beberapa tahapan yang harus dilakukan untuk membentuk sebuah pohon keputusan dengan menerapkan algoritma C4.5 adalah sebagai berikut (Galih, 2019):

1. Menyiapkan data latih

Data latih merupakan data histori yang sudah pernah diolah sebelumnya dan sudah dibagi menjadi beberapa kelas yang telah ditentukan.

2. Menentukan akar dari pohon

Dalam menentukan akar diperoleh dari atribut yang telah terpilih, yang dilakukan dengan proses menghitung nilai gain yang ada pada setiap atribut dan jika nilai gain paling tinggi yang dihasilkan maka akan menjadi akar pertama. Sebelum menentukan nilai gain, ditentukan terlebih dahulu nilai entropy yang merupakan distribusi probabilitas pada teori informasi yang digunakan oleh Pohon Keputusan (C4.5) dalam menentukan tingkat kesamaan (homogenitas) distribusi kelas dari sebuah dataset. Jika pada sebuah dataset memiliki nilai yang tinggi tingkat entropy nya maka semakin homogen distribusi kelas tersebut.

Untuk menghitung nilai entropy digunakan rumus :

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

Keterangan :

- S = himpunan (dataset) kasus
- N = banyaknya partisi S
- Pi = proporsi Si terhadap S

3. Menghitung nilai Gain dengan persamaan

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

Keterangan :

- S = Himpunan Kasus
- A = Fitur
- n = Jumlah Partisi Atribut A
- |Si| = Proporsi Si terhadap S
- |S| = Jumlah Kasus dalam S

Beberapa tahapan yang harus dilakukan untuk melakukan seleksi fitur menggunakan *Adaboost Classification*, sebagai berikut :

Tahap seleksi fitur pada penelitian ini dilakukan menggunakan metode *Adaboost Classification*, dengan cara kerjanya yang memberikan bobot yang sama pada setiap sampel awalnya, lalu setelah setiap diklasifikasi bobot yang benar berkurang dan bobot hasil yang salah bertambah maka dilakukan proses ini berulang hingga mencapai batas atau jumlah siklus maksimum

5. Evaluasi

Evaluasi merupakan suatu tahapan yang digunakan untuk mengetahui kinerja sistem, salah satunya dengan dilakukannya evaluasi berupa pengecekan akurasi.

$$Akurasi = \frac{TP + TN}{TP + TN + FP + FN}$$

Pada confusion matrik tersebut ada 4 kolom yaitu (Daqiqil, 2021) :

1. TP (True Positif) berisi jumlah data points diberi label Yes yang memang sebenarnya bernilai Yes.
2. TN (True Negatif) berisi jumlah data points diberi label No yang memang sebenarnya bernilai No.
3. FP (False Positif) berisi jumlah data points diberi label Yes yang memang sebenarnya bernilai No. Error ini biasa disebut Error Type 1.
4. FN (False Negatif) berisi jumlah data points diberi label No yang memang sebenarnya bernilai Yes. Error ini biasa disebut Error Type 2.

Hasil Dan Pembahasan

Business Understanding (Pemahaman Bisnis)

Penelitian ini menggunakan nilai akhir mahasiswa yang diperoleh dari bagian Mata Kuliah Dasar Umum (MKDU) UMKT dan Bagian Administrasi Akademik (BAA) UMKT. Data nilai mata kuliah yang digunakan yaitu Bahasa Indonesia seluruh program studi angkatan 2020 dan 2021 dengan records data sebanyak 1028 mahasiswa.

Data Understanding (Pemahaman Data)

Data nilai akhir mahasiswa yang diperoleh dari BAA memiliki 5 atribut diantaranya, jenis kelamin, nomor induk mahasiswa (nim), nama, nilai akhir, bobot dan simbol.

Tabel 1 Data BAA

No	Atribut	Keterangan
1	NIM	Nomor induk mahasiswa
2	Nama Mahasiswa	Nama mahasiswa
3	Nilai Akhir	Nilai akhir mahasiswa
4	Bobot	Bobot nilai standar penilaian
5	Simbol	Skala penilaian
6	Jenis Kelamin	Jenis Kelamin mahasiswa

Data nilai mahasiswa yang diperoleh dari MKDU pada mata kuliah “Bahasa Indonesia” memiliki 18 atribut diantaranya, *profile name, learner name, learner email, enrolment ID, institution membership id, enrolment date, completion date, time spent on course, progress, %course completed, certificate ID, comments, kudos, enrolment cost, tugas1, tugas2, tugas3, dan UTS.*

Table 2 Data MKDU

Data Preparation

Data Selection dan Integration

Dengan adanya tahap seleksi dan integrasi, atribut yang tidak diperlukan seperti nim, nama, dan nilai akhir dari data BAA dihapuskan. Sementara itu atribut yang dihapuskan pada data MKDU adalah profile name, learner name, learner email, enrolment ID, institution

No	Atribut	Keterangan
1	Profile name	Id mahasiswa pada sistem OpenLearning
2	Leamer name	Nama mahasiswa
3	Leamer email	Email mahasiswa
4	Enrolment ID	Id pendaftaran OpenLearning
5	Institution Membership ID	Id anggota institusi
6	Enrolment date	Tanggal daftar
7	Completion date	Tanggal menyelesaikan mata kuliah
8	Time spent on course	Lama waktu mahasiswa berada di mata kuliah
9	Progress	Persentase kemajuan mahasiswa
10	%Course Completed	Persentase kemajuan mahasiswa dalam menyelesaikan mata kuliah
11	Certificate ID	ID sertifikat
12	Comments	Banyaknya komentar mahasiswa selama perkuliahan
13	Kudos	Penghargaan
14	Enrolment Cost	Biaya pendaftaran
15	Tugas 1	Nilai mahasiswa pada tugas yang pertama dimata kuliah
16	Tugas 2	Nilai mahasiswa pada tugas yang kedua dimata kuliah
17	Tugas 3	Nilai mahasiswa pada tugas yang ketiga dimata kuliah
18	UTS	Nilai mahasiswa ujian yang dilakukan tengah semester dimatakuliah

membership id, enrolment date, completion date, certificate ID, kudos, enrolment cost. Pada tabel dibawah merupakan hasil dari tahap integrasi data MKDU dan data BAA menjadi satu dan seleksi atribut pada data.

Table 3 Data Setelah Seleksi dan Integrasi

No	Jenis Kelamin	Time spent on course	Progress	% Course completed	Comments	Tugas 1	Tugas 2	Tugas 3	UTS	Simbol
1	Perempuan	10 Hrs 43 Mins	99,37	99,37	61	88	80	80	77	B
2	Laki-laki	10 Hrs 45 Mins	100	100	139	82	80	82	80	A
3	Laki-laki	4 Hrs 57 Mins	100	100	38	84	80	80	77	A
4	Perempuan	7 Hrs 25 Mins	100	100	86	82	80	84	78	A
5	Perempuan	12 Hrs 0 Mins	100	100	29	84	82	84	75	A
:	:	:	:	:	:	:	:	:	:	:
1028	Perempuan	10 Hrs 57 Mins	96,86	96,86	14	80	80	80	90	AB

Transformasi data

Tahap transformasi data dilakukan guna mengubah struktur, format, nilai ataupun tipe pada dataset sesuai dengan kebutuhan. Ada 3 atribut data yang melakukan perubahan tipe data yaitu time spent on course, jenis kelamin dan symbol.

Pada atribut "time spent on course", awalnya direpresentasikan dalam format waktu seperti "jam:menit:detik", maka dapat diubah menjadi tipe data numerik seperti jumlah total detik atau menit.

Tabel 4 Transformasi Data Atribut Time Spent On Course

Time spent on course (Sebelum Transformasi)	Time spent on course (Sesudah Transformasi)
10 Hrs 43 Mins	643
10 Hrs 45 Mins	645
4 Hrs 57 Mins	297
7 Hrs 25 Mins	445
12 Hrs 0 Mins	720
:	:
10 Hrs 57 Mins	657

Berdasarkan ketentuan standar kelulusan mata kuliah jika nilai mahasiswa A, AB, atau B maka ditransformasi menjadi Lulus Lulus sebagai kelas "1", dan mahasiswa yang bernilai BC, C, D dan E maka kelas diubah menjadi Tidak Lulus sebagai kelas "0".

Table 5 Transformasi Data Atribut Simbol

Simbol (Sebelum Transformasi)	Simbol (Sesudah Transformasi)	Simbol(Sesudah Transformasi sebagai kelas 1 dan 0)
B	LULUS	1
A	LULUS	1

A	LULUS	1
A	LULUS	1
A	LULUS	1
:	:	:
AB	LULUS	1

Tahap transformasi terakhir dilakukan pada atribut jenis kelamin yang semula direpresentasikan sebagai "Laki-laki" menjadi "1" dan "Perempuan" menjadi "2".

Table 6 Transformasi Atribut Jenis Kelamin

Jenis Kelamin(Sebelum Transformasi)	Jenis Kelamin(Sesudah Transformasi)
Perempuan	2
Laki-laki	1
Laki-laki	1
Perempuan	2
Perempuan	2
:	:
Perempuan	2

Reduksi Data

Tahap ini melakukan pembagian kelas yang tidak balance serta menghindari data yang tidak konsisten. Dalam penelitian ini kelas target Baik sebanyak 969 dan Buruk sebanyak 59 lalu diterapkan teknik undersampling yaitu yang bertujuan untuk mengurangi jumlah sampel dari kelas target mayoritas dengan jumlah records data 59 yang diambil secara acak.

```
Distribusi kelas setelah reduksi:
1 59
0 59
Name: Simbol, dtype: int64
```

Gambar 3 Reduksi Data

Berikut tabel yang telah melalui tahap reduksi dan pembersihan data

Table 7 Hasil Data Reduksi

No	Jenis Kelamin	Time spent on course	Progress	% Course completed	Comments	Tugas 1	Tugas 2	Tugas 3	UTS	Simbol
1	1	490	100.0	100.0	23	90	85	80	46	1
2	1	338	100.0	100.0	24	64	73	73	65	1
3	2	4080	100.0	100.0	24	76	78	78	61	1
4	2	468	100.0	100.0	25	69	72	63	41	1
5	1	294	100.0	100.0	71	86	78	80	77	1
:	:	:	:	:	:	:	:	:	:	:
118	1	236	62.89	62.89	10	78	67	78	85	0

Modelling (Pemodelan)

Pemodelan Algoritma C4.5 Dengan Python

Setelah itu melewati beberapa tahap data preparation sebelumnya, dataset kini siap untuk dilakukan pengujian dengan menerapkan algoritma C4.5.

```
Fold-1 Confusion Matrix:      Fold-6 Confusion Matrix:
[[5 1]                        [[5 1]
 [0 6]]                        [1 5]]
Accuracy: 0.85                 Accuracy: 0.67

Fold-2 Confusion Matrix:      Fold-7 Confusion Matrix:
[[5 1]                        [[2 4]]
 [0 6]]                        [2 4]]
Accuracy: 0.85                 Accuracy: 0.67

Fold-3 Confusion Matrix:      Fold-8 Confusion Matrix:
[[5 1]                        [[2 4]]
 [0 6]]                        [2 4]]
Accuracy: 0.85                 Accuracy: 0.67
```

Hasil evaluasi kinerja algoritma C4.5 rata-rata confusion matrix dan akurasi dapat dilihat pada Gambar 4 dan penjabaran confusion matrix pada Tabel 7 dengan akurasi pada perhitungan dibawah ini

```
Average Confusion Matrix:
[[4.8 1.1]
 [1.2 4.7]]
Average Accuracy: 0.81
```

Gambar 2 Tampilan Cross Validation dari Fold 1 sampai 10

```
Fold-5 Confusion Matrix:      Fold-10 Confusion Matrix:
[[6 0]                         [[5 1]
 [2 4]]                         [4 7]]
Accuracy: 0.85                 Accuracy: 0.81
```

Gambar 4 Avarage Confusion Matrix

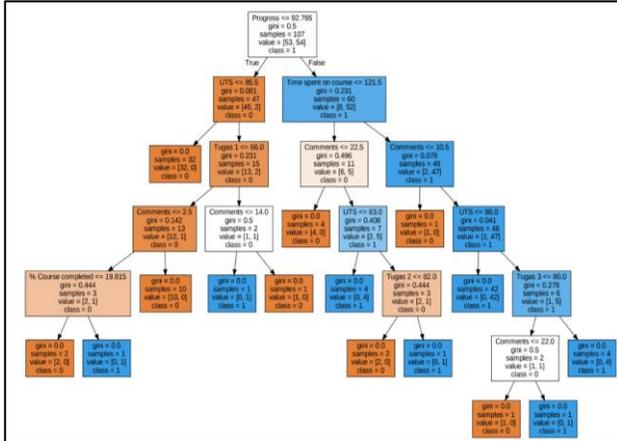
Table 8 Confusion Matrix

Actual	Predict	
	Lulus	Tidak Lulus
Lulus	48	11
Tidak Lulus	12	47

$$\begin{aligned}
 Accuracy &= \frac{TP + TN}{TP + TN + FP + FN} \times 100 \% \\
 &= \frac{48 + 47}{48 + 47 + 11 + 12} \times 100 \% \\
 &= \frac{95}{118} \times 100 \% \\
 &= 0.80508474 \text{ atau } 81 \%
 \end{aligned}$$

Pada tingkat akar (node 0), fitur yang digunakan untuk membagi sampel adalah "Progress" yang memiliki nilai ambang batas <= 92.765. Sampel dengan nilai "Progress" di bawah ambang batas akan diklasifikasikan ke kelas 0

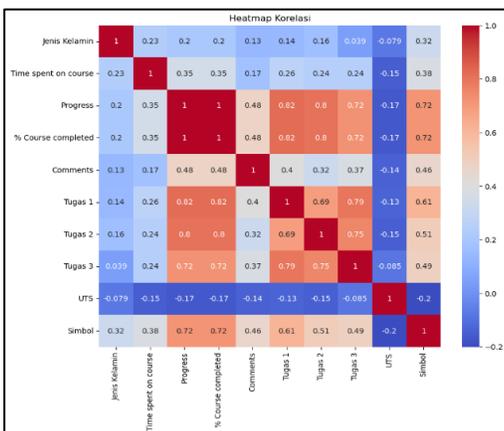
(Tidak Lulus), sedangkan sampel dengan nilai "Progress" di atas atau sama dengan ambang batas akan diklasifikasikan ke kelas I (Lulus).



Gambar 5 Hasil Pohon Keputusan C4.5

Pemodelan Algoritma C4.5 Dan Metode Adaboost Classifier

Tahap ini dilakukan untuk menyeleksi fitur dengan menggunakan Adaboost Classifier guna melihat korelasi antara atribut jenis kelamin, time spent on course, progress, % course completed, comments, tugas 1, tugas 2, tugas 3, terhadap simbol sebagai kelas target. Hasil dari seleksi fitur dengan menambahkan metode adaboost didalamnya didapatkan atribut yang memiliki korelasi pada simbol dengan nilai >0.50 yaitu progress, % course completed, tugas 1, dan tugas 2.



Pada tahap selanjutnya pengujian ini menggunakan dataset dengan atribut yang sudah terseleksi dengan metode Adaboost Classifier yang memiliki korelasi terhadap simbol dengan melakukan pemodelan dengan C4.5

dengan tujuan dapat mengoptimalkan akurasi pada algoritma tersebut.

Fold-1 Confusion Matrix: [[5 1] [0 6]] Accuracy: 0.92	Fold-6 Confusion Matrix: [[6 0] [0 6]] Accuracy: 1.00
Fold-2 Confusion Matrix: [[5 1] [1 5]] Accuracy: 0.83	Fold-7 Confusion Matrix: [[4 2] [2 4]] Accuracy: 0.67
Fold-3 Confusion Matrix: [[5 1] [1 5]] Accuracy: 0.83	Fold-8 Confusion Matrix: [[6 0] [0 6]] Accuracy: 1.00
Fold-4 Confusion Matrix: [[6 0] [2 4]] Accuracy: 0.83	Fold-9 Confusion Matrix: [[5 0] [3 3]] Accuracy: 0.73
Fold-5 Confusion Matrix: [[6 0] [0 6]] Accuracy: 1.00	Fold-10 Confusion Matrix: [[5 1] [0 5]] Accuracy: 0.91

Gambar 7 Urutan Fold 1-10

Hasil evaluasi berupa accuracy dengan teknik confusion matrix yang mampu memberikan hasil yang optimal pada algoritma C4.5 menggunakan metode Adaboost Classifier.

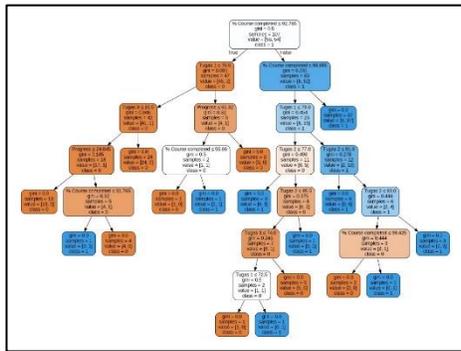
Average Confusion Matrix:
[[5.1 0.8]
[1. 4.9]]
Average Accuracy: 0.85

Gambar 8 Avarage Confusion Matrix

Pada pengujian pemodelan algoritma C4.5 dan metode Adaboost Classification didapatkan hasil rata-rata akurasi dari semua fold dalam 10-fold cross-validation, dengan perhitungan akurasi dibawah ini sebesar 85% yang mana menunjukkan bahwa model berhasil meningkatkan akurasi sebesar 4%, dari 81% menjadi 85%.

$$\begin{aligned}
 & \text{Accuracy} \\
 &= \frac{TP + TN}{TP + TN + FP + FN} \times 100 \% \\
 &= \frac{51 + 49}{51 + 49 + 8 + 10} \times 100 \% \\
 &= \frac{100}{118} \times 100 \% \\
 &= 0.847457 \text{ atau } 85 \%
 \end{aligned}$$

Hasil dari visualisasi pohon keputusan, pada tingkat akar (node 0), fitur yang digunakan untuk membagi sampel adalah "Progress" yang memiliki nilai ambang batas <= 92.765. Sampel dengan nilai "Progress" di bawah ambang batas akan diklasifikasikan ke kelas 0 (Tidak Lulus), sedangkan sampel dengan nilai "Progress" di atas atau sama dengan ambang batas akan diklasifikasikan ke kelas I (Lulus).



Gambar 9 Hasil Pohon Keputusan

Simpulan Dan Saran

Dalam implementasi Algoritma C4.5 dan Seleksi Fitur Adaboost Classification, metode K-Fold Cross Validation dengan skema 10-Fold digunakan untuk pembagian data, karena menghasilkan akurasi yang konsisten. Implementasi Algoritma C4.5 pada klasifikasi nilai Mahasiswa Mata Kuliah Bahasa Indonesia Tahun Akademik 2020/2021 menggunakan data dari Bagian Administrasi & Akademik dan Bagian Mata Kuliah Dasar Umum Universitas Muhammadiyah Kalimantan Timur. Tingkat akurasi yang diperoleh adalah 81% dan 85%. Terdapat peningkatan akurasi sebesar 5% ketika Seleksi Fitur Adaboost Classification diterapkan pada Algoritma C4.5. Rekomendasi penambahan atribut termasuk absensi kehadiran, informasi keuangan mahasiswa, penghasilan orang tua, dan profesi. Dengan meningkatkan jumlah atribut, diharapkan akurasi klasifikasi dapat ditingkatkan. Penelitian selanjutnya dapat mempertimbangkan penggunaan metode seleksi fitur alternatif seperti Forward Selection, Backward Elimination, dan metode lainnya pada algoritma C4.5. Selain itu, eksplorasi penggunaan Seleksi Fitur Adaboost Classification dengan algoritma machine learning yang berbeda juga disarankan. Tujuannya adalah menggabungkan teknik seleksi fitur dan algoritma machine learning untuk mencapai hasil yang optimal.

Pustaka Acuan

- Anggita, S. D., & Ikmah, I. (2021). Implementasi Pso Untuk Optimalisasi Bobot Atribut Pada Algoritma C4.5 Dalam Prediksi Kelulusan Mahasiswa. *JUPI (Jurnal Ilmiah Penelitian Dan Pembelajaran Informatika)*, 6(2), 416–423. <https://doi.org/10.29100/jupi.v6i2.2440>
- Atma, Y. D., & Setyanto, A. (2018). Perbandingan algoritma c4.5 dan k-nn dalam identifikasi mahasiswa berpotensi drop out. *Metik Jurnal ISSN : 2580-1503*, 2(2), 31–37.
- Galih, G. (2019). Data Mining di Bidang Pendidikan untuk Analisa Prediksi Kinerja Mahasiswa dengan Komparasi 2 Model Klasifikasi pada STMIK Jabar. *Jurnal Teknologi Sistem Informasi Dan Aplikasi*, 2(1), 23. <https://doi.org/10.32493/jtsi.v2i1.2643>
- Komang Aditya Pratama, Gede Aditra Pradnyana, & I Ketut Resika Arthana. (2020). Pengembangan Sistem Cerdas Untuk Prediksi Daftar Kembali Mahasiswa Baru Dengan Metode Naive Bayes (Studi Kasus: Universitas Pendidikan Ganesha). *SINTECH (Science and Information Technology) Journal*, 3(1), 22–34. <https://doi.org/10.31598/sintechjournal.v3i1.523>
- Rustam, S. (2020). Klasifikasi Kompetensi Mahasiswa Dengan Algoritma Decision Tree Dalam Menentukan Kelayakan Mata Kuliah Kosentrasi. *Simtek: Jurnal Sistem Informasi Dan Teknik Komputer*, 5(2), 59–62. <https://doi.org/10.51876/simtek.v5i2.76>
- Wanto, A., Siregar, N. H., Windarto, A. P., Hartama, D., Ginantra, N. L. W. S. R., Napitulu, D., Negara, E. S., Lubis, M. R., Dewi, S. V., & Prianto, C. (2020). *Data Mining ALGORITMA & IMPLEMENTASI* (T. Limbong (ed.)). Yayasan Kita Menulis.