

Peningkatan Algoritma J48 Untuk Klasifikasi Hasil Prestasi Mahasiswa Selama Proses Pembelajaran Secara Daring Menggunakan CFS dan Adaboost

Dian Pramadhana ^{#1}, Rendi ^{#2}, Robi Robiyanto ^{#3}

^{#1,2,3} *Teknik Informatika, ^{#1,2,3} Politeknik Negeri Indramayu
Jl. Raya Lohbener Lama No. 8 Indramayu 45252*

¹ dianpramadhana@polindra.ac.id

² rendi@polindra.ac.id

³ robiyanto@polindra.ac.id

Accepted on 12-12-2022

Abstrak

Virus corona yaitu virus yang menyebabkan infeksi saluran pernapasan pada manusia. Jumlah kenaikan penderita virus corona meningkat secara cepat dan menyebar diberbagai negara. Dalam menghadapi virus corona dunia pendidikan mengantisipasi agar penularan virus corona tidak semakin luas yaitu dengan menghentikan proses pembelajaran secara langsung dan digantikan dengan pembelajaran daring. Dalam proses pembelajaran daring tidak selalu berjalan secara baik ataupun lancar, banyak kendala yang dihadapi selama pembelajaran daring. Kendala yang ada selama proses pembelajaran daring berakibat terhadap perubahan peserta didik seperti menurunnya motivasi serta prestasi belajar mahasiswa. Berdasarkan alasan tersebut peneliti tertarik untuk menganalisis hasil prestasi mahasiswa selama pembelajaran secara daring dengan menggunakan teknik *data mining* yaitu teknik klasifikasi. Salah satu teknik *data mining* untuk melakukan klasifikasi yang dapat digunakan adalah algoritma J48. Algoritma J48 merupakan algoritma yang cukup populer dalam pengklasifikasian, namun memiliki kelemahan yaitu dalam proses klasifikasi tidak dapat memilih atribut yang relevan, hal ini dapat berdampak buruk saat proses membangun *decision tree*. Untuk mengatasi masalah tersebut peneliti menerapkan *Correlation Feature Selection* untuk seleksi atribut dan Adaboost dalam memperbaiki performa algoritma J48. Hal ini dibuktikan bahwa setelah penerapan metode yang digunakan hasil akurasi yang di peroleh menunjukkan peningkatan yaitu sebesar 96,42 %.

Keywords: *AdaBoost*, Algoritma *J48*, *CFS*, Data Mining, Klasifikasi, Virus Corona.

I. INTRODUCTION

Virus corona yaitu virus yang dapat menyebabkan infeksi saluran pernapasan pada manusia [1]. Jumlah kenaikan penderita virus corona meningkat secara cepat. Virus corona menyebar di berbagai negara sehingga WHO menjadikan wabah virus corona sebagai pandemi. Kasus virus corona terjadi pertama di Indonesia dikonfirmasi pada tanggal 2 Maret 2020 [2]. Pemerintah melakukan upaya pencegahan penularan virus corona dengan menerapkan beberapa aturan, yaitu dengan menggunakan masker saat diluar rumah, sering mencuci tangan, menjaga jarak dan melarang adanya kerumunan. Dengan adanya kasus virus corona yang semakin meningkat, lembaga pendidikan menyampaikan surat edaran Kemendikbud Dikti No.

1 tahun 2020 [3] dengan memerintahkan perguruan tinggi tidak melakukan pembelajaran secara tatap muka dan diganti dengan pembelajaran secara daring [4]. Selama proses pembelajaran daring, mahasiswa dan dosen bisa menggunakan aplikasi Whatsapp atau aplikasi sejenisnya [5]. Dalam proses pembelajaran daring, tidak selamanya berjalan secara baik ataupun lancar. Terdapat beberapa kendala yang ditemukan selama proses pembelajaran secara daring.

Kendala yang ada selama proses pembelajaran daring berakibat terhadap perubahan peserta didik seperti menurunnya motivasi serta prestasi belajar. Motivasi pada mahasiswa serta prestasi belajar adalah 2 aspek yang saling berkaitan. Hasil prestasi belajar peserta didik akan menjadi baik apabila peserta didik memiliki motivasi yang tinggi dalam belajar, sehingga apabila peserta didik memiliki motivasi yang kurang dalam belajar maka akan berpengaruh terhadap hasil prestasi [6]. Dari kasus ini peneliti memanfaatkan teknik *data mining* yaitu klasifikasi dalam mengetahui hasil prestasi belajar mahasiswa selama proses pembelajaran secara daring.

Data mining adalah salah satu cara mendapatkan pola yang menarik serta pengetahuan pada sebuah data dengan jumlah yang banyak [7]. Beberapa teknik dapat digunakan dalam *data mining* yaitu dengan pengenalan pola, melakukan prediksi, klasifikasi, asosiasi serta *clustering* [8]. Klasifikasi pada *data mining* adalah salah satu cara dalam menemukan sekumpulan model dengan menjelaskan atau membedakan sebuah kelas, dengan tujuan agar bisa memprediksi suatu kelas pada objek yang kelasnya belum diketahui sebelumnya [9]. Pada dasarnya, sebuah algoritma klasifikasi dalam proses untuk menggunakan semua fitur yang ada di dalam data yang digunakan untuk membangun sebuah model. Padahal, fitur yang ada dalam data tersebut tidak selalu semua relevan terhadap hasil klasifikasi yang diinginkan. Sehingga, apabila hal tersebut terjadi pada sebuah data yang mempunyai ukuran yang besar, maka akan berakibat kinerja sebuah algoritma menjadi kurang efektif dan efisien. Sebagai contoh dengan melakukan proses yang memerlukan waktu yang lama untuk pengklasifikasian. Karena harus memproses banyak fitur tersebut. Proses seleksi fitur yaitu dengan melakukan pemilihan fitur yang relevan pada data dan akan menghilangkan fitur yang dianggap tidak relevan atau redundan yang dapat mempengaruhi terhadap hasil klasifikasi [10].

Selection feature adalah aspek penting yang dapat digunakan dalam melakukan pengoptimalan kinerja dari *classifier*. Dalam menggunakan *selection feature* yang sesuai pada saat melakukan proses klasifikasi akan meningkatkan hasil akurasi yang didapat [11]. Cara kerja dari seleksi fitur yaitu fitur yang dianggap kurang relevan atau redundan akan dihilangkan agar proses pengklasifikasian bisa lebih baik. Beberapa kasus yang terjadi *selection feature* bisa meningkatkan kinerja dari sebuah algoritma. Dalam meningkatkan kinerja *machine learning* perlu dilakukan pemilahan subset fitur yang dianggap relevan pada sekelompok data besar, hal ini perlu dilakukan dalam mengoptimalkan kinerja *machine learning* [12]. Beberapa metode yang bisa dipakai pada seleksi fitur yaitu metode *Correlation Feature Selection*. Sebuah fitur yang kurang relevan dapat mengakibatkan kinerja *machine learning* menjadi berkurang. Kemudian, pada fitur yang redundan berakibat kinerja *machine learning* memerlukan waktu yang lama dalam pemrosesan. Memilih subset fitur yang relevan pada data merupakan hal yang penting dilakukan karena akan meningkatkan kinerja dari *machine learning* [13].

Algoritma J48 adalah sebuah pengembangan dari sebuah algoritma konvensional induksi pohon keputusan yaitu ID3. Algoritma J48 adalah algoritma yang paling populer digunakan untuk pengklasifikasian. Ada beberapa metode yang dapat meningkatkan akurasi pengklasifikasian, diantaranya adalah metode *ensemble* yaitu *Bagging* dan *Boosting* merupakan metode *ensemble* yang dapat meningkatkan akurasi klasifikasi. AdaBoost dipergunakan dalam mengurangi kesalahan yang terjadi pada algoritma pembelajaran dan terbukti dengan signifikan dapat memperoleh hasil klasifikasi yang lebih optimal. Algoritma Adaboost akan memberi bobot di seluruh data latih, bobot tersebut akan menentukan distribusi data untuk pembangunan model [14].

II. LITERATURE REVIEW

A. Prestasi Belajar

Prestasi Belajar merupakan hasil nilai dari perumusan akhir yang diberikan seorang guru yang berkaitan dengan kemajuan hasil prestasi belajar seorang siswa selama proses pembelajaran [15]. Prestasi

belajar adalah alat yang digunakan dalam melakukan pengukuran terhadap siswa sejauh mana telah mencapai tujuan yang sudah ditetapkan [16].

B. Pembelajaran daring

Proses pembelajaran secara daring adalah metode pembelajaran dengan menggunakan dan memanfaatkan jaringan internet pada saat proses pembelajaran. Proses pembelajaran secara daring dapat diartikan juga sebagai proses pendidikan yang dilaksanakan oleh pihak sekolah yang peserta didiknya dan gurunya berada di tempat yang berbeda dimana membutuhkan alat telekomunikasi interaktif sebagai media untuk penghubung dan beberapa sumber daya yang dibutuhkan di dalamnya [17].

C. Data Mining

Data mining adalah salah satu cara menemukan pola yang menarik, serta pengetahuan pada data dengan jumlah data yang sangat besar. *Data mining* merupakan metode dengan proses pencarian yang melalui data berjumlah besar, dengan usaha agar menemukan sebuah tren, pola serta sebuah hubungan. [18]. *Data mining* merupakan suatu cara dalam memperoleh kecerdasan pada sebuah data berjumlah besar.

Dari definisi tersebut, dapat diambil kesimpulan yaitu *data mining* merupakan proses menganalisis dalam mendapatkan sebuah informasi yang belum diketahui dengan menggunakan statistik serta *artificial intelligence* pada suatu *database* yang memiliki jumlah sangat besar, sehingga didapatkan sebuah pola dari sebuah data yang tidak diketahui sebelumnya dan pola itu selanjutnya diwujudkan dalam bentuk grafik komputer sehingga mudah dimengerti.

D. Classification

Classification merupakan sebuah bentuk menganalisis sebuah data yang menghasilkan sebuah model untuk menggambarkan sebuah kelas pada data yang penting. Proses *classification* akan melakukan prediksi kategori (*discrete, unordered*) terhadap label *class*. *Classification* adalah sebuah proses untuk memperoleh sebuah model dan fungsi yang menjelaskan dan membedakan sebuah kelas data, yang memiliki tujuan agar bisa memperkirakan sebuah kelas dari suatu objek yang belum diketahui labelnya [19].

E. Correlation Feature Selection (CFS)

CFS adalah salah satu algoritma filter sederhana dengan proses kerjanya melakukan perangkingan dari sebuah subset yang sesuai dengan fungsi evaluasi heuristik berbasis korelasi. Dimana suatu subset atribut yang dianggap baik adalah atribut yang mempunyai kolerasi yang tinggi terhadap kelasnya serta tidak mempunyai korelasi dengan atribut lainnya. Dimana sebuah atribut yang mempunyai korelasi dengan atribut lain maka atribut tersebut merupakan atribut yang redundan. Sedangkan sebuah atribut yang memiliki korelasi rendah terhadap kelas, maka atribut itu merupakan atribut yang tidak relevan. Sebuah atribut yang tidak relevan serta redundan maka harus dihilangkan [20].

F. Algoritma J48

Algoritma *J48* adalah sebuah algoritma yang menggunakan pohon keputusan. Metode pohon keputusan akan melakukan perubahan fakta yang sangat besar menjadi pohon keputusan yang merepresentasikan menggunakan aturan. Dimana aturan bisa secara mudah dimengerti dengan bahasa alami. Pohon keputusan dapat juga diekspresikan pada bentuk bahasa basis yaitu *Structured Query Language* untuk menemukan *record* pada kategori tertentu [21].

G. AdaBoost

Boosting terbukti bisa meningkatkan sebuah kinerja dari pengklasifikasi pada banyak situasi, salah satunya saat data tidak seimbang [22]. *Boosting* merupakan sebuah metode yang secara umum dapat mengoptimalkan kerja sebuah algoritma. *Boosting* dapat diterapkan dalam memperkecil terjadinya kesalahan dari sebuah algoritma.

III. RESEARCH METHOD

A. Metode Penelitian

Penelitian dilakukan dengan menerapkan metode data mining untuk menentukan faktor penyebab menurunnya hasil prestasi mahasiswa selama masa pandemi di Politeknik Baja Tegal. Dalam penelitian ini menggunakan metode yaitu *Correlation Feature Selection*, metode ini digunakan sebagai seleksi atribut dan Algoritma *J48* digunakan untuk proses klasifikasi. Agar dalam pengklasifikasian kinerja algoritma *J48* meningkat, maka ditambahkan *Adaboost*. *Adaboost* digunakan dalam meningkatkan hasil akurasi yang didapat dengan cara memberikan bobot terhadap setiap atribut, sehingga kinerja algoritma *J48* semakin baik.

B. Rancangan Penelitian

Beberapa tahapan dalam penelitian ini meliputi identifikasi masalah dan kebutuhan, analisis dan perencanaan, uji coba dan evaluasi, serta pelaporan hasil penelitian.

1. Identifikasi Masalah

Identifikasi masalah merupakan proses menganalisis sebuah permasalahan yang sedang terjadi atau hal yang sedang akan diteliti. Dalam penelitian ini identifikasi masalah terkait dengan penentuan faktor penyebab menurunnya hasil prestasi mahasiswa di masa pandemi.

2. Pengumpulan Data

Pengumpulan data pada penelitian ini penulis menggunakan angket kuisioner sebagai sarana dalam pengumpulan data terkait hasil pembelajaran daring. Pengisian angket dilakukan secara daring dengan menggunakan Google Form. Data yang digunakan pada penelitian ini sebanyak 84 data. Data tersebut selanjutnya dilakukan analisis dengan cara analisis deskriptif.

3. Studi Pustaka

Studi pustaka adalah cara yang dilakukan dalam melengkapi pengetahuan dasar akan teori-teori yang digunakan pada penelitian ini.

4. Pengujian Metode yang diusulkan

Pada penelitian ini untuk melakukan percobaan serta pengujian pada data, peneliti memanfaatkan *tools WEKA*. Pada tahap pertama dilakukan proses seleksi atribut terhadap data dengan *Corelation Feature Selection* dimana ini bertujuan untuk menghilangkan atribut yang dianggap kurang mendukung serta redundan. Selanjutnya dilakukan proses klasifikasi dengan menggunakan algoritma *J48* serta penambahan *Adaboost* yang bertujuan untuk mengurangi terjadinya kesalahan dari algoritma *J48*. Sehingga hasil kinerja pengklasifikasiannya dapat lebih optimal.

5. Evaluasi dan Validasi

Pada penelitian ini untuk mengetahui tingkat akurasi dari metode yang diusulkan menggunakan evaluasi *confusion matrix* untuk mengetahui nilai akurasi, *precision*, serta *recall*. Evaluasi ini bertujuan untuk menilai berapa persen kinerja metode yang digunakan untuk pengklasifikasian data prestasi mahasiswa.

IV. RESULTS AND DISCUSSION

A. Data yang digunakan

Data yang digunakan adalah data kuisioner yang sudah diberikan pada mahasiswa Politeknik Baja Tegal dengan jumlah 84 data. Data kemudian diolah memakai *software WEKA* untuk memperoleh hasil dari metode yang digunakan. Berikut data yang digunakan ditampilkan pada Tabel I.

TABEL I
 DATA PRESTASI MAHASISWA

No	A1 (Media)	A2 (Durasi)	A3 (Lingkungan)	A4 (Jaringan)	A5 (Penjelasan)	A6 (Pemahaman)	A7 (Tugas)	A8 (Waktu)	A9 (Minat)	Hasil
1	HP	Sebentar	Tidak_Kondusif	Baik	Jelas	Cukup	Jarang	Tepat	Menarik	Menurun
2	HP	Sebentar	Kondusif	Cukup	Jelas	Cukup	Sering	Tidak_Tepat	Menarik	Meningkat
3	Laptop	Lama	Kondusif	Kurang	Jelas	Kurang	Sering	Tepat	Menarik	Meningkat
4	HP	Sebentar	Kondusif	Cukup	Cukup	Cukup	Sering	Tepat	Menarik	Meningkat
5	HP	Sebentar	Kondusif	Cukup	Kurang	Cukup	Sering	Tepat	Tidak_Menarik	Menurun
::	::	::	::	::	::	::	::	::	::	::
83	HP	Sebentar	Tidak_Kondusif	Cukup	Kurang	Cukup	Jarang	Tepat	Tidak_Menarik	Menurun
84	HP	Sebentar	Tidak_Kondusif	Cukup	Kurang	Cukup	Jarang	Tepat	Tidak_Menarik	Menurun

B. Penerapan Metode CFS (Correlation Feature Selection)

Tahap awal penelitian dilakukan pengolahan data memakai *Correlation Feature Selection*. Sebelum melakukan perhitungan menggunakan *Correlation Feature Selection*, setiap dari atribut dilakukan perhitungan korelasi atribut dengan kelas serta atribut dengan atribut yang lainnya. *Correlation Feature Selection* akan memilih atribut relevan serta tidak redundan dengan melihat nilai merit yang tertinggi. *Correlation Feature Selection* akan mencari subset yang semula atribut 0. Setelah itu, ditambahkan 1 atribut dan dihitung nilai merit. Selanjutnya, dihitung sampai semua atribut diuji coba. Nilai merit atribut yang paling tinggi akan dipilih dimana subset ini merupakan hasil dari *Correlation Feature Selection*. Tabel II ini merupakan hasil dari rangkuman nilai merit tertinggi yang didapatkan setelah melakukan perhitungan dengan menggunakan rumus dari *Correlation Feature Selection*.

TABEL II
 RANGKUMAN HASIL NILAI MERIT PALING TINGGI

No	Attribute	Merit
1	A.7	0,833
2	A.7, A.6	1,209
3	A.7, A.6, A.5	1,348
4	A.7, A.6, A.5, A.1	1,427
5	A.7, A.6, A.5, A.1, A.9	1,436
6	A.7, A.6, A.5, A.1, A.9, A.3	1,471
7	A.7, A.6, A.5, A.1, A.9, A.3, A.8	1,467
8	A.7, A.6, A.5, A.1, A.9, A.3, A.8, A.4	1,297
9	A.7, A.6, A.5, A.1, A.9, A.3, A.8, A.4, A.2	1,117

Dari Tabel II nilai merit yang tertinggi didapatkan dari nomor 6 adalah atribut 7, 6, 5, 1, 9 dan 3 dengan nilai merit yaitu sebesar 1,471. Hasil menunjukkan bahwa atribut 7, 6, 5, 1, 9 dan 3 relevan pada kelas serta tidak redundan pada atribut yang lain. Jadi, atribut 2, atribut 4, dan atribut 8 dapat dihilangkan dikarenakan atribut 2, atribut 4, dan atribut 8 redundan serta tidak relevan. Tabel III merupakan data prestasi mahasiswa setelah dilakukan proses *Correlation Feature Selection*.

TABEL III
 DATA SETELAH DILAKUKAN PROSES CFS

No	A1 (Media)	A3 (Lingkungan)	A5 (Penjelasan)	A6 (Pemahaman)	A7 (Tugas)	A9 (Minat)	Hasil
1	HP	Tidak_Kondusif	Jelas	Cukup	Jarang	Menarik	Menurun
2	HP	Kondusif	Jelas	Cukup	Sering	Menarik	Meningkat
3	Laptop	Kondusif	Jelas	Kurang	Sering	Menarik	Meningkat
4	HP	Kondusif	Cukup	Cukup	Sering	Menarik	Meningkat
5	HP	Kondusif	Kurang	Cukup	Sering	Tidak_Menarik	Menurun
::	::	::	::	::	::	::	::
83	HP	Tidak_Kondusif	Kurang	Cukup	Jarang	Tidak_Menarik	Menurun
84	HP	Tidak_Kondusif	Kurang	Cukup	Jarang	Tidak_Menarik	Menurun

C. Metode J48

Tahap berikutnya setelah dilakukan proses *Correlation Feature Selection* kemudian mengklasifikasikan menggunakan *J48* sesuai atribut yang terpilih. Tabel IV merupakan tabel hasil dari kinerja algoritma *J48* sebelum dan sesudah menggunakan *Correlation Feature Selection* di data prestasi mahasiswa sebagai berikut:

TABEL IV
 HASIL PENGUKURAN J48 TANPA MENGGUNAKAN CORRELATION FEATURE SELECTION

Pengukuran	Hasil
<i>Accuracy</i>	92.85 %
<i>Recall</i>	0.929
<i>Precision</i>	0.931
<i>AUC</i>	0.926

Hasil yang didapat yaitu sebesar 92.85 % untuk *accuracy*, 0.929 untuk *recall*, 0.931 untuk *precision* dan nilai *AUC* diperoleh yaitu 0.926. Lihat hasilnya pada Tabel V.

TABEL V
 HASIL PENGUKURAN J48 MENGGUNAKAN CORRELATION FEATURE SELECTION

Pengukuran	Hasil
<i>Accuracy</i>	93.11 %
<i>Recall</i>	0.931
<i>Precision</i>	0.933
<i>AUC</i>	0.929

Hasil yang didapat yaitu sebesar 93.11 % untuk *accuracy*, 0.931 untuk *recall*, 0.933 untuk *precision* dan nilai *AUC* diperoleh yaitu 0.929.

D. J48 Berbasis Adaboost

Tahap akhir untuk mengoptimalkan kinerja *J48* dengan menambahkan *AdaBoost*. Hasil untuk uji memakai data prestasi mahasiswa yang telah diterapkan proses *Correlation Feature Selection* memakai algoritma *J48* berbasis *AdaBoost* memperlihatkan kenaikan nilai hasil *accuracy* yaitu 96.42 %, 0.964 untuk nilai *recall*, 0.965 untuk nilai *precision* dan 0.943 untuk *Area Under Curve (AUC)*. Dari hasil tersebut memperlihatkan metode yang dipakai dapat meningkatkan hasil dari akurasi terhadap data prestasi mahasiswa. Tabel VI merupakan hasil pengukuran dari metode yang digunakan.

TABEL VI
 HASIL ACCURACY, RECALL, PRECISION, DAN AUC

Pengukuran	Hasil
<i>Accuracy</i>	96.42 %
<i>Recall</i>	0.964
<i>Precision</i>	0.965
<i>AUC</i>	0.943

E. Evaluasi dan Validasi Hasil

Pada bagian ini merupakan hasil dari evaluasi dan validasi hasil yang menggunakan *Confusion Matrix*. Untuk hasil data yang lengkap, dapat dilihat pada Tabel VII.

TABEL VII
 HASIL CONFUSION MATRIX

	MENURUN	MENINGKAT
MENURUN	25 (TP)	1 (FN)
MENINGKAT	2 (FP)	56 (TN)

Dibawah ini merupakan hasil perhitungan dari *confusion matrix* berdasarkan dari Tabel VII:

1) *Accuracy* merupakan hasil dari prediksi yang betul, rumus untuk menghitung *Accuracy* sbb:

$$\begin{aligned}
 Accuracy &= \frac{TP + TN}{TP + TN + FP + FN} \\
 &= \frac{25 + 56}{25 + 56 + 2 + 1} \\
 &= \frac{81}{84} = 0,9642
 \end{aligned}
 \tag{1}$$

2) *Sensitivity (recall) / TP rate* merupakan kasus yang diamati bernilai positif, dihitung menggunakan rumus dibawah ini:

$$\begin{aligned}
 Recall &= \frac{TP}{TP + FN} \\
 &= \frac{25}{25 + 1} = \frac{25}{26} = 0,962
 \end{aligned}
 \tag{2}$$

3) *Precision* adalah tingkat kesesuaian informasi yang diinginkan pengguna terhadap jawaban yang diberikan sistem, rumus perhitungan sbb :

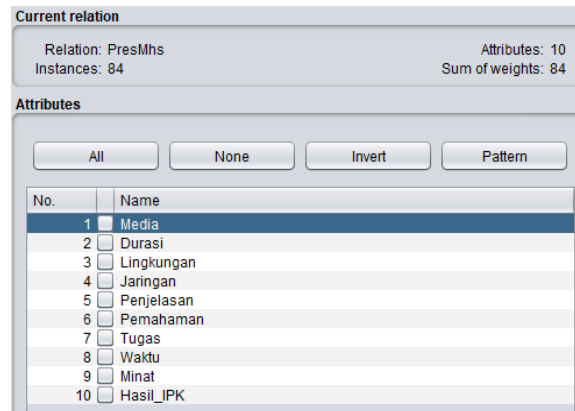
$$\begin{aligned}
 Precision\ Positive &= \frac{TP}{TP + FP} \\
 &= \frac{25}{25 + 2} = \frac{25}{27} = 0,926
 \end{aligned}
 \tag{3}$$

4) *AUC* memperlihatkan seberapa benar dan sukses sebuah model yang dipergunakan dengan melakukan pemisahan nilai negatif serta positif, rumus perhitungan sbb:

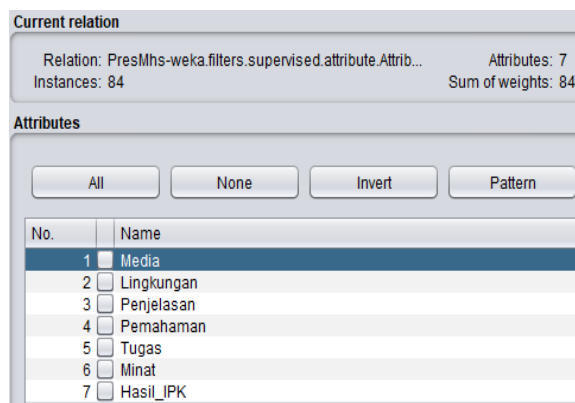
$$\begin{aligned}
 AUC &= \frac{1 + TP_{rate} - FP_{rate}}{2} \\
 &= \frac{1 + 0.962 - 0.034}{2} \\
 &= 0,943
 \end{aligned}
 \tag{4}$$

F. Pembahasan

Dari Pengujian metode yang digunakan yaitu dengan menerapkan metode *Correlation feature selection* terhadap data Prestasi Mahasiswa dengan jumlah 84 data. Kemudian algoritma *J48* yang berbasis *AdaBoost* digunakan untuk pengklasifikasian dengan tahapan mula-mula dilakukan proses awal menggunakan *Correlation-based Feature Selection*. Hal ini berguna untuk menghilangkan atribut yang redundant serta tidak relevan terhadap dataset prestasi mahasiswa. Gambar 1 merupakan gambar hasil dari seleksi atribut sebelum diproses *Correlation feature selection*. Sedangkan pada Gambar 2 merupakan gambar hasil dari seleksi atribut setelah diproses *Correlation feature selection*



Gambar 1. Atribut sebelum proses *Correlation Feature Selection*

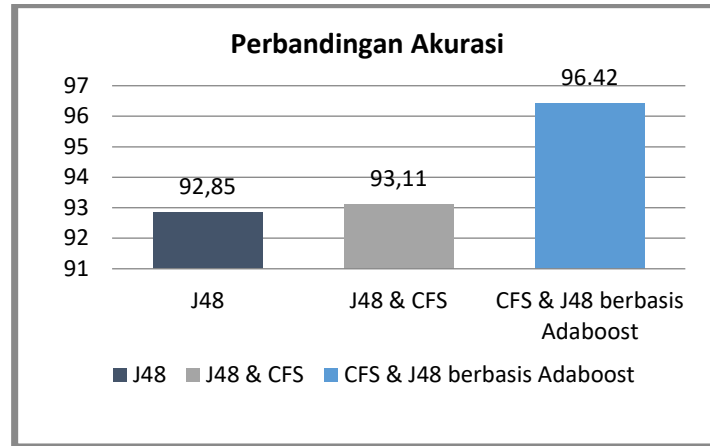


Gambar 2. Atribut setelah proses *Correlation Feature Selection*

Pada data awal dengan jumlah 9 atribut yaitu Media, Lingkungan, Jaringan, Penjelasan, Pemahaman, Tugas, Waktu dan Minat diperoleh 6 atribut dari 9 atribut. Dari 6 atribut tersebut memperlihatkan atribut ini relevan pada kelas serta tidak redundant pada atribut lainnya. Hasil dari 6 atribut terpilih tersebut adalah Media, Lingkungan, Penjelasan, Pemahaman, Tugas dan Minat. Sedangkan atribut Durasi, Jaringan dan Waktu tidak digunakan karena atribut tersebut tidak relevan.

Kemudian proses selanjutnya memakai algoritma *J48* berbasis *AdaBoost* untuk mengklasifikasikan. Dari hasil eksperimen menggunakan *AdaBoost* menunjukkan pengoptimalan bobot yang menjadikan nilai

bobot naik. Ketika sebelum menerapkan *AdaBoost* dan *Correlation Feature Selection*, diperoleh hasil akurasi sebesar 92,85%. Setelah menerapkan *Correlation Feature Selection*, akurasi yang diperoleh menjadi sebesar 93,11%. Setelah Penerapan *Correlation Feature Selection* dan *AdaBoost* menunjukkan peningkatan hasil akurasi yaitu 96,42%. Pemangkasan atribut oleh *Correlation Feature Selection* serta perbaikan bobot dari *AdaBoost* mengurangi kesalahan pengklasifikasian dari algoritma *J48*. Dari hasil tersebut, Gambar 3 adalah hasil dari perbandingan hasil akurasi sebelum dan sesudah diterapkannya metode yang digunakan.



Gambar 3. Hasil Perbandingan Akurasi

Tabel VIII adalah hasil akhir dari nilai yang didapatkan dengan menerapkan *Correlation Feature Selection* pada algoritma *J48* berbasis *Adaboost* menggunakan tools *Weka*.

TABEL VIII
 HASIL CORRELATION FEATURE SELECTION PADA ALGORITMA J48 BERBASIS ADABOOST

Pengukuran	Hasil
<i>Accuracy</i>	96.42 %
<i>TP Rate</i>	0,926
<i>FP Rate</i>	0,342
<i>Precision</i>	0.965
<i>Recall</i>	0.964
<i>AUC</i>	0.943

V. Conclusion

Berdasarkan penelitian yang sudah dilaksanakan didapatkan hasil penggunaan seleksi fitur menggunakan metode *Corretion Feature Selection* serta penambahan *AdaBoost* menunjukkan peningkatan hasil akurasi yang diperoleh algoritma *J48* yaitu sebesar 96,42%. Dengan penggunaan *Correlation Feature Selection* dan *AdaBoost* dalam melakukan klasifikasi bisa meningkatkan kinerja dari Algoritma *J48*, sehingga kinerja dari algoritma *J48* semakin baik dibandingkan tanpa menggunakan metode *Correlation Feature Selection* dan *AdaBoost*. Penelitian yang sudah dilakukan tersebut, menunjukkan hasil nilai akurasi lebih optimal dan lebih baik. Saran untuk penelitian selanjutnya diharapkan dapat meningkatkan performa kinerja algoritma *J48* sehingga hasil akurasi yang diperoleh jauh lebih baik dari penelitian ini.

REFERENCES

- [1] H. A. Rothan and S. N. Byrareddy, "The epidemiology and pathogenesis of coronavirus disease (COVID-19) outbreak," *J. Autoimmun.*, vol. 109, no. February, p. 102433, 2020, doi: 10.1016/j.jaut.2020.102433.
- [2] R. Tosepu *et al.*, "Correlation between weather and Covid-19 pandemic in Jakarta, Indonesia," *Sci. Total Environ.*, vol. 725, p. 138436, 2020, doi: 10.1016/j.scitotenv.2020.138436.

- [3] UNIVERSITAS ANDALAS, “Kementerian Pendidikan dan Kebudayaan,” [Http://Kemdikbud.Go.Id/](http://Kemdikbud.Go.Id/), vol. 126, no. 021, pp. 1–3, 2020, [Daring]. Available: <http://kemdikbud.go.id/main/?lang=id>
- [4] K. F. Irnanda, D. Hartama, and A. P. Windarto, “Analisa Klasifikasi C4.5 Terhadap Faktor Penyebab Menurunnya Prestasi Belajar Mahasiswa Pada Masa Pandemi,” *J. Media Inform. Budidarma*, vol. 5, no. 1, p. 327, 2021, doi: 10.30865/mib.v5i1.2763.
- [5] A. Cahyani, I. D. Listiana, and S. P. D. Larasati, “Motivasi Belajar Siswa SMA pada Pembelajaran Daring di Masa Pandemi Covid-19,” *IQ (Ilmu Al-qur’an) J. Pendidik. Islam*, vol. 3, no. 01, pp. 123–140, 2020, doi: 10.37542/iq.v3i01.57.
- [6] S. Agarwal, *Data mining: Data mining concepts and techniques*. 2014. doi: 10.1109/ICMIRA.2013.45.
- [7] F. Gorunescu, *Data Mining Concepts, Models and Techniques*. Springer ed, 2011.
- [8] V. Kumar and N. Rathee, “Knowledge discovery from database using an integration of clustering and classification,” *Int. J. Adv. Comput. Sci. Appl.*, vol. 2, no. 3, pp. 29–33, 2011, doi: 10.14569/ijacsa.2011.020306.
- [9] B. K. Khotimah, U. Trunojoyo, and J. Timur, “Seleksi Fitur Dan Parameter Optimal Dengan,” pp. 1–11, 2012.
- [10] A. A. Saifan and L. A. Abuwardih, “Software defect prediction based on feature subset selection and ensemble classification,” *ECTI Trans. Comput. Inf. Technol.*, vol. 14, no. 2, pp. 213–228, 2020, doi: 10.37936/ecti-cit.2020142.224489.
- [11] P. Thanathamathee and C. Lursinsap, *Handling imbalanced data sets with synthetic boundary data generation using bootstrap re-sampling and AdaBoost techniques*, vol. 34, no. 12. Elsevier B.V., 2013. doi: 10.1016/j.patrec.2013.04.019.
- [12] P. Yildirim, “Filter Based Feature Selection Methods for Prediction of Risks in Hepatitis Disease,” *Int. J. Mach. Learn. Comput.*, vol. 5, no. 4, pp. 258–263, 2015, doi: 10.7763/ijmlc.2015.v5.517.
- [13] W. Liu, S. Chawla, D. A. Cieslak, and N. V. Chawla, “A robust decision tree algorithm for imbalanced data sets,” *Proc. 10th SIAM Int. Conf. Data Mining, SDM 2010*, no. June 2014, pp. 766–777, 2010, doi: 10.1137/1.9781611972801.67.
- [14] D. Pramadhana, “Klasifikasi Penyakit Diabetes Menggunakan Metode CFS dan ROS dengan Algoritma J48 Berbasis Adaboost,” *Edumatic J. Pendidik. Inform.*, vol. 5, no. 1, pp. 89–98, 2021, doi: 10.29408/edumatic.v5i1.3336.
- [15] Faishal, “Efektivitas pembelajaran bahasa via daring,” *Ta’dibi J. Manaj. Pendidik. Islam*, vol. IX, no. September 2020, pp. 114–140, 2021, [Daring]. Available: <http://e-jurnal.stail.ac.id/index.php/tadibi/article/view/226/144>
- [16] Rosyid, Mustajab, Abdullah, 2019. *Prestasi Belajar*. Malang : CV. Literasi Nusantara.
- [17] Mustofa, MI, dkk. (2019). Formulasi Model Perkuliahan Daring Sebagai Upaya Menekan Disparitas Kualitas Perguruan Tinggi. *WJIT: Walisongo Journal of Information Tecnology*, 1 (2), 151-160.
- [18] Gorunescu. F. *Data Mining: Concepts, Models and Techniques*, Berlin, Springer-Verlag. 2011.
- [19] Pascual. D. *Artificial Intelligence Tools*, CRC Pres Taylor&Prancis Group. 2015.
- [20] Afza. A. et al. *A Hybrid Classifier using Boosting, Clustering, and Naïve Bayesian Classifier*, World of Computer Science and Information Technology Journal (WCSIT). 105-109, 2018.
- [21] W. Liu, S. et al. *A robust decision tree algorithm for imbalanced data sets*, Conference on Data Mining. pp. 1 – 12, 2016.
- [22] Afza. A. et al. *A Hybrid Classifier using Boosting, Clustering, and Naïve Bayesian Classifier*, World of Computer Science and Information Technology Journal (WCSIT). 105-109, 2017.