

Implementasi Fase *Boosting* pada Algoritma C5.0 dalam Menentukan Prestasi Akademik Siswa

Salsabila Citra Esananda¹, Budi Nugroho^{2*}, Fetty Tri Anggraeny³

^{1,3} Informatika, Universitas Pembangunan Nasional “Veteran” Jawa Timur)

¹citrasce@gmail.com

³fettyanggraeny.if@upnjatim.ac.id

² Informatika, Universitas Pembangunan Nasional “Veteran” Jawa Timur)

*Corresponding author email: budinugroho.if@upnjatim.ac.id

Abstrak— Prestasi akademik siswa merupakan hasil dari kegiatan pembelajaran yang disertai perubahan atas pencapaian siswa dan dinyatakan dalam bentuk huruf, angka, maupun kalimat sebagai ukuran keberhasilan dengan standar yang telah ditetapkan. Penentuan prestasi akademik siswa dilakukan secara manual oleh tenaga pendidikan, hal ini sangat tidak efektif pada waktu yang dibutuhkan mengingat semakin banyaknya kuantitas siswa dan rentan akan *human error*. Untuk itu, dalam penelitian ini penulis menerapkan algoritma *decision tree* C5.0 yang mana terdapat fase *boosting* untuk mengoptimalkan model pohon dalam mengklasifikasikan prestasi akademik siswa dengan memperbarui data *training* untuk membentuk model pohon keputusan yang paling akurat. Penelitian ini menghasilkan model pohon keputusan dengan akurasi 93,15%. Kemudian, dilakukan pengujian sistem menggunakan 144 data *testing* yang menghasilkan nilai *recall* sebesar 84,32%, *precision* sebesar 81,62%, dan nilai *accuracy* sebesar 84,03%.

Kata Kunci— Prestasi, Klasifikasi, Algoritma C5.0, *Boosting*, *Confusion Matrix*.

I. PENDAHULUAN

Prestasi akademik adalah sebuah proses yang dialami siswa dan menghasilkan perubahan dalam bidang pengetahuan, pemahaman, penerapan, daya analisis, sintesis dan evaluasi[1]. Prestasi akademik siswa dinyatakan dalam bentuk simbol, angka, huruf, maupun kalimat sebagai tingkat ukuran keberhasilan dengan standar yang telah ditetapkan. Prestasi belajar siswa dipengaruhi oleh dua faktor, yaitu faktor internal yang berasal dari dalam diri sendiri seperti fisiologis dan psikologis siswa itu sendiri. Untuk faktor eksternal yang berasal dari luar yang dipengaruhi oleh lingkungan keluarga, sekolah, masyarakat, maupun alam [2].

Penggunaan teknologi informasi dalam pengolahan data termasuk dalam menentukan prestasi akademik siswa sangatlah membantu. Hal ini akan membuat lebih efektif dan efisien dalam *monitoring* perkembangan siswa dibandingkan dengan cara manual yang rentan akan *human error* serta memakan banyak waktu mengingat data yang semakin hari semakin banyak. Namun hingga saat ini belum dijumpai instansi pendidikan yang menggunakan sistem dalam mengklasifikasikan prestasi akademik siswa.

Oleh karena itu, melalui penelitian ini penulis membangun sistem penentuan prestasi akademik siswa dengan menerapkan teori *data mining*, salah satunya dengan teknik klasifikasi.

Beberapa algoritma klasifikasi dalam *data mining*, diantaranya seperti *Decision Tree Induction*, *Bayesian Classification*, *Support Vector Machine Classification*, *Neural Network Classification*, dsb. Sehingga dalam melakukan klasifikasi penentuan prestasi akademik siswa dapat dengan menerapkan metode *decision tree* baik itu menggunakan algoritma ID3, C4.5 ataupun C5.0.

Algoritma C5.0 merupakan salah satu metode klasifikasi dari *Decision Tree* sehingga menerapkan struktur pohon atau struktur berhierarki dalam mengambil keputusan. Algoritma C5.0 merupakan pengembangan dari metode *Decision Tree* yang sebelumnya, yaitu algoritma ID3 dan C4.5. Sehingga kinerja algoritma ini lebih unggul dari algoritma sebelumnya dengan adanya fase *boosting* di proses terakhir dalam perhitungan algoritma C5.0.

Dengan adanya fase tersebut, akurasi pohon keputusan yang dihasilkan algoritma C5.0 lebih optimal karena selalu dilakukan pembaruan bobot pada data *training* yang digunakan sebagai acuan untuk memperbarui data *training* dalam membentuk pohon keputusan. Ketika data *training* diperbarui kemudian diolah menggunakan perhitungan algoritma *decision tree*, maka akan menghasilkan model pohon keputusan yang baru juga. Dari beberapa proses pembaruan data *training* ini, akan menghasilkan beberapa model pohon keputusan yang pada tahap akhir akan dipilih salah satu model pohon yang lebih efektif daripada pohon lainnya untuk dijadikan sebagai acuan dalam pengambilan keputusan dalam sistem yang dibangun.

II. METODOLOGI PENELITIAN

A. Data Mining

Data mining merupakan cara untuk menggambarkan pola dan kecenderungan yang dimiliki data [3]. Secara umum terdapat lima peranan *data mining*, diantaranya estimasi, prediksi, klasifikasi, *clustering*, dan asosiasi. Proses pengolahan data dalam *data mining* memerlukan sebuah algoritma yang dikelompokkan berdasarkan masing-masing peranan, pada kasus estimasi dan prediksi banyak digunakan algoritma *Linear Regression*, *Support Vector Machine*, *Neural Network*, sedangkan untuk klasifikasi banyak yang menggunakan algoritma *K-Nearest Neighbor*, *Naive Bayes*, *Decision Tree* (ID3, C4.5, C5.0, CART). Dan dalam kasus *clustering*, dapat digunakan algoritma *K-Means*, *Fuzzy*, *C-Means*, *K-Medoid*, *Self-Organization Map*. Untuk asosiasi

digunakan algoritma *FP-Growth*, *A priori*, *Chi Square*, dan *Coefficient of Correlation* [4].

B. Algoritma C5.0

Algoritma C5.0 merupakan salah satu algoritma klasifikasi *data mining* penerapan dari teknik *Decision Tree* yang merupakan penyempurnaan dari algoritma sebelumnya yaitu ID3 dan C4.5, dibentuk oleh John Ross Quinlan pada tahun 1987. Dalam proses pembentukan pohon keputusan pada algoritma ini, didasarkan pada atribut yang memiliki nilai informasi *gain* tertinggi akan dijadikan sebagai *root* bagi *node* selanjutnya lalu atribut yang dipilih akan menjadi pembagi untuk sampel tersebut [5].

Algoritma C5.0 menghasilkan pohon dengan jumlah cabang *node* yang bervariasi sehingga diperlukan strategi pembentukan *decision tree* dengan menggunakan algoritma C5.0 sebagai berikut [6] :

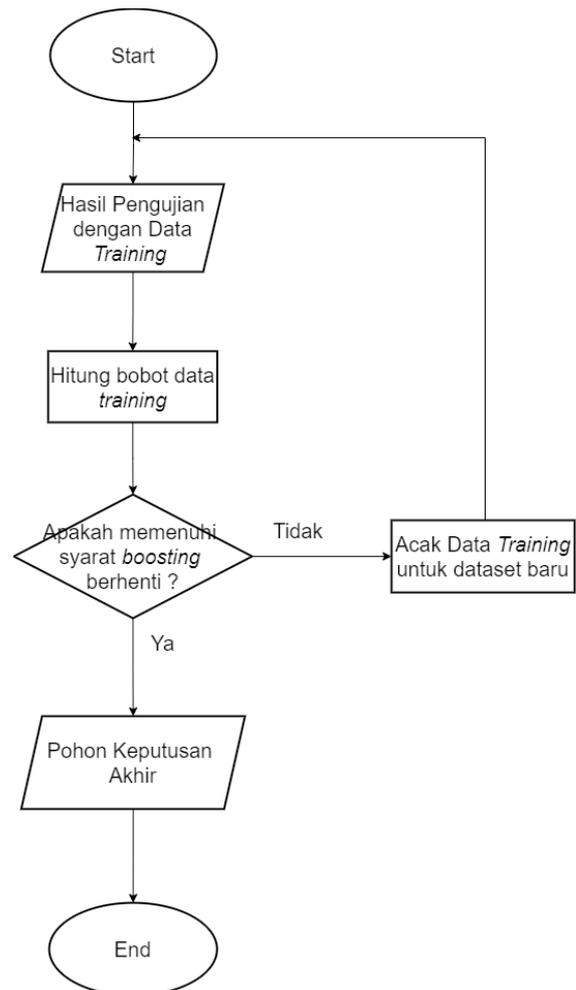
1. Proses pertama *tree* digambarkan sebagai *node* tunggal yang merepresentasikan *training* set.
2. Jika sampel seluruhnya berada pada kelas yang sama (terjadi kondisi *confidence value*), maka *node* tersebut menjadi *leaf* dan dilabeli dengan kelas tersebut.
3. Jika terdapat sampel data di kelas lain, algoritma dengan menggunakan ukuran berbasis *entropy* (*information gain*) akan memilih variabel prediktor yang akan memisahkan *record* ke dalam kelas-kelas individual sehingga variabel tersebut menjadi variabel tes atau keputusan pada *node* tersebut.
4. Cabang dikembangkan untuk setiap nilai yang diketahui dari variabel tes, dan sampel dipartisi berdasarkan cabang tersebut.
5. Algoritma menggunakan proses yang sama secara rekursif membentuk *decision tree*.
6. Partisi rekursif berakhir ketika terjadi kondisi-kondisi sebagai berikut :
 - a. Seluruh *record* pada *node* tertentu memiliki kelas yang sama.
 - b. Tidak ada atribut yang tersisa pada *record* yang dapat dipartisi lebih lanjut. Dalam kasus ini suara mayoritas digunakan. *Node* tersebut menjadi *leaf node* dan dilabeli dengan kelas yang menjadi mayoritas dalam *record* yang ada.
 - c. Tidak ada *record* untuk cabang variabel tes. Dalam kasus ini, *leaf node* terbentuk dengan mayoritas kelas sebagai label *record* tersebut.

Dari beberapa proses algoritma C5.0, pembentukan pohon keputusan diawali dengan memilih salah satu atribut data untuk dijadikan sebagai *node* pertama (*root*) dari pohon keputusan, yang didasarkan pada nilai *information gain* tertinggi dari atribut yang lain. Algoritma ini membentuk pohon keputusan dengan cara pembagian dan menguasai sampel secara rekursif dari atas ke bawah. Algoritma ini dimulai dengan semua data yang dijadikan akar dari pohon keputusan sedangkan atribut yang dipilih menjadi pembagi bagi sampel tersebut.

C. Boosting

Boosting merupakan sebuah teknik untuk meningkatkan akurasi prediksi dengan mengoptimalkan kinerja model pohon atau mesin klasifikasi yang pada akhirnya digunakan sebagai

acuan dalam pengujian data dengan cara memberikan dan menghitung bobot pada data *training* [7]. Berikut ini diagram alir dari fase *boosting* pada algoritma C5.0.



Gbr 1. Diagram Alir Fase *Boosting*

Boosting dilakukan secara berulang dengan mengacak data *training* sebelumnya untuk membentuk data *training* baru secara random yang selanjutnya data tersebut diolah dalam perhitungan algoritma C5.0 agar terbentuk model pohon keputusan baru sesuai dengan data *training* yang telah diacak. Sehingga setiap dilakukan perulangan (*iterasi*) *boosting* otomatis akan menghasilkan bobot baru yang berbeda dengan bobot pada data sebelumnya. Proses ini dimulai dengan memberikan bobot yang sama pada setiap data dengan menggunakan persamaan rumus dibawah ini.

$$w = \frac{1}{N} \quad (1)$$

Keterangan :

w = bobot

N = jumlah data yang digunakan untuk membentuk model (data *training*)

Bobot data disini akan berubah setelah dilakukan beberapa tahap proses perhitungan. Setelah mendapatkan nilai bobot awal, untuk meng-*update* bobot diperlukan menghitung nilai titik tengah (*midpoint*) antara yang jumlah data yang salah

diklasifikasi dan data yang benar diklasifikasi secara keseluruhan. Untuk nilai benar klasifikasi dan salah klasifikasi didapatkan dari hasil pengujian data *training* menggunakan pohon keputusan yang telah terbentuk. Rumus untuk menghitung *midpoint* dengan menggunakan persamaan 2 dibawah ini.

$$midpoint = \frac{1}{4}(S_+ - S_-) \quad (2)$$

Keterangan :

S_- = jumlah bobot data yang salah diklasifikasi

S_+ = jumlah bobot data yang benar diklasifikasi

Setelah diperoleh nilai *midpoint*, kemudian melakukan perhitungan untuk mendapatkan nilai bobot baru pada data yang terklasifikasikan dengan benar menggunakan persamaan 3 ini.

$$w_k = w_{k-1} \times \frac{S_+ - midpoint}{S_+} \quad (3)$$

Dan untuk mendapatkan nilai bobot baru pada data yang salah klasifikasi dapat menggunakan persamaan 4 dibawah ini.

$$w_k = w_{k-1} + \frac{midpoint}{N_-} \quad (4)$$

Keterangan :

w_k = jumlah bobot data pada iterasi *boosting* ke- k

w_{k-1} = jumlah bobot data pada iterasi *boosting* ke- $k-1$

N_- = jumlah data yang salah klasifikasi

Setelah selesai dilakukan perhitungan bobot menggunakan beberapa persamaan di atas, dilakukan iterasi yang pertama dengan mengacak data *training* awal untuk membentuk data *training* baru dan selanjutnya dilakukan perhitungan bobot yang sama seperti pada proses di atas.

Proses perulangan dalam mengacak data *training* dapat berhenti dilakukan setelah dilakukan beberapa iterasi pengacakan data atau ketika model pohon keputusan yang terbentuk sudah dinilai efektif, seperti pada kondisi jumlah bobot data yang salah diklasifikasikan (S_-) bernilai kurang dari 0,10 atau rata-rata dari bobot data yang salah diklasifikasikan bernilai lebih dari 0,5 [8]. Setelah selesai dilakukan beberapa iterasi pada proses ini, kemudian dari beberapa model pohon keputusan yang terbentuk dilakukan perbandingan model pohon keputusan yang paling efektif atau memiliki akurasi tertinggi untuk dijadikan sebagai model pohon keputusan akhir yang digunakan sebagai acuan dalam pengambilan keputusan pada tahap pengujian sistem menggunakan data *testing* yang telah disediakan.

D. Confusion Matrix

Confusion matrix adalah metode yang digunakan untuk mengukur performa dari *machine learning*. Terdapat empat istilah yang digunakan untuk merepresentasikan hasil proses klasifikasi data terhadap model pohon keputusan, diantaranya adalah : True Positive (TP), False Positive (FP), True Negative (TN), dan False Negative (FN) [9]. Keempat nilai tersebut dibagi pada tabel di bawah ini.

TABEL 1
CONFUSION MATRIX

Aktual	Kelas	Prediksi	
		A	B
	A	TP	FP
	B	FN	TN

Dari tabel confusion matrix diatas, dilakukan perhitungan dari hasil pengujian sistem untuk menghasilkan 3 nilai, diantaranya nilai *recall*, *precision*, dan *accuracy* model menggunakan persamaan di bawah ini.

$$recall = \frac{TP}{TP+FN} * 100\% \quad (5)$$

$$precision = \frac{TP}{TP+FP} * 100\% \quad (6)$$

$$accuracy = \frac{TP+TN}{TP+TN+FP+FN} * 100\% \quad (7)$$

III. HASIL DAN PEMBAHASAN

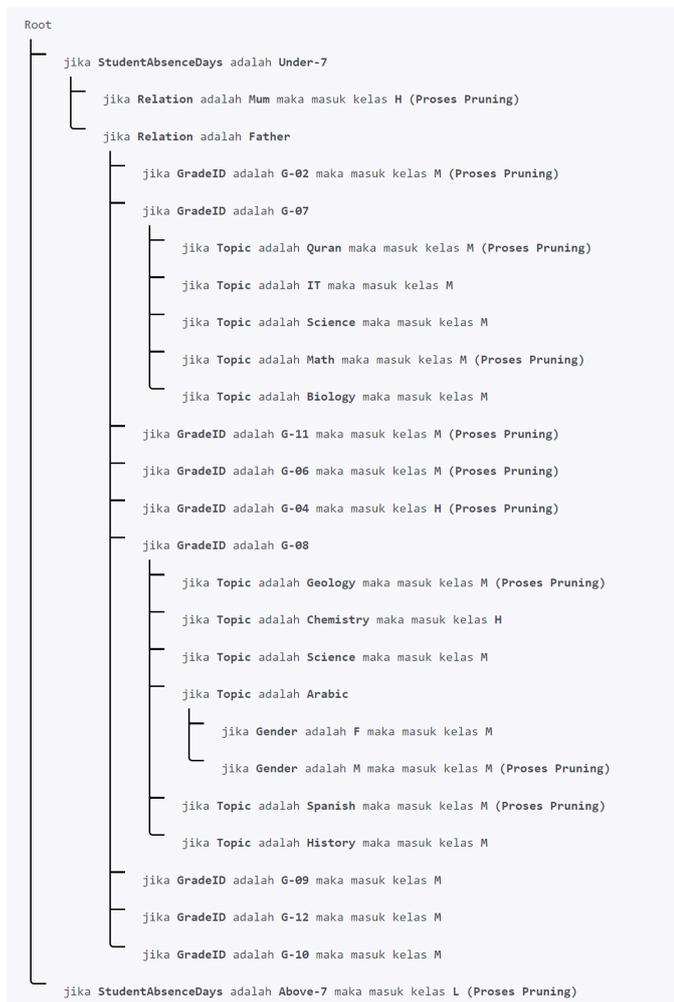
A. Implementasi Data

Dataset penelitian yang digunakan diambil dari website *kaggle.com* (*Student's Academic Performance Dataset*), dengan tujuan untuk mengklasifikasikan prestasi akademik siswa. Dataset ini merupakan kumpulan data siswa dengan jumlah 480 data yang diambil dari sistem manajemen pembelajaran *Kalboard360* melalui xAPI, digunakan dalam penelitian sebelumnya yang dilakukan Elaf Abu Amrieh, Thair Hamtini, dan Ibrahim Aljarah [10]. Data ini terbagi ke dalam 3 kelas, yaitu kelas H, M, dan L serta memiliki 16 atribut yang seluruhnya digunakan dalam penelitian ini diantaranya, *gender, nationality, place of birth, stage id, grade id, section id, topic, semester, relation, raisedhands, visited resources, announcements view, discussion, parents answering survey, parents school satisfaction, dan students absence day*.

Dari keseluruhan dataset tersebut, dilakukan pembagian data dengan proporsi 70:30. Sebanyak 336 data sebagai data *training* dan sisanya 144 data sebagai data *testing*.

B. Pembentukan Model

Pada tahap ini dilakukan pengolahan pada 336 data *training* sesuai dengan perhitungan pada algoritma C5.0 untuk membentuk suatu model pohon keputusan. Dari perhitungan tersebut menghasilkan model pohon keputusan sebagai berikut :



Gbr 2. Pohon Keputusan Awal

Pada pengolahan pertama menghasilkan model pohon keputusan seperti gambar di atas, dimana dari 336 data *training* sistem mampu memprediksi data dengan benar sejumlah 231 data, sedangkan sebanyak 105 salah diklasifikasikan. Dari hasil pengujian tersebut dilakukan perhitungan bobot dalam fase *boosting* untuk menentukan nilai bobot awal dengan persamaan 1 sebagai berikut.

$$w_0 = \frac{1}{N} = \frac{1}{336} = 0,002976$$

Setelah diketahui nilai bobot awal, substitusikan ke dalam persamaan 2 untuk mencari nilai *midpoint*.

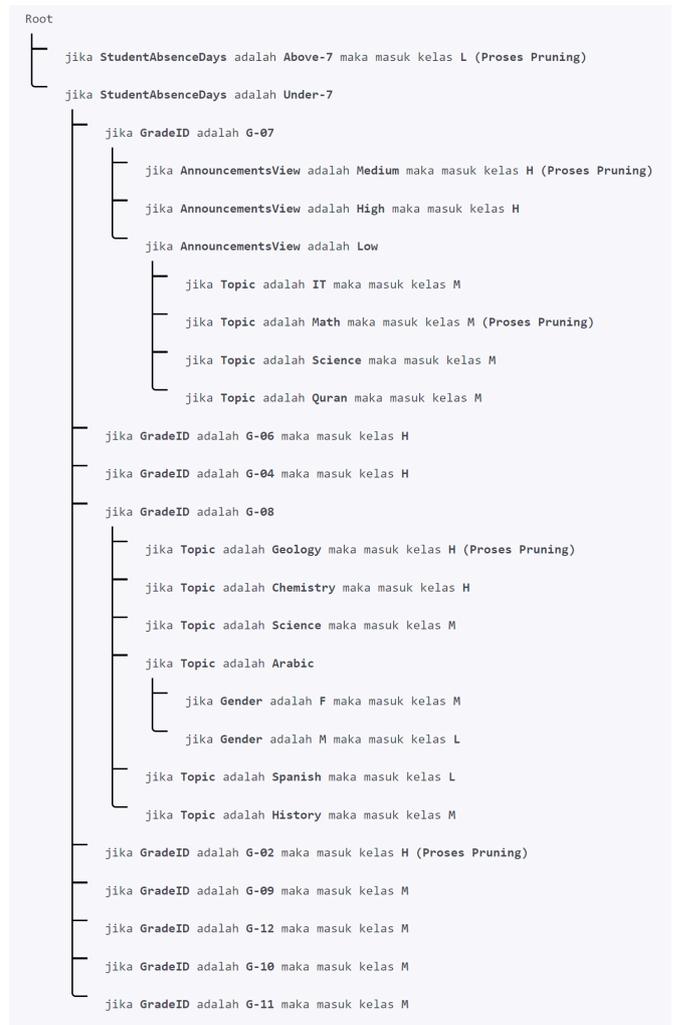
$$\begin{aligned} \text{midpoint} &= \frac{1}{4}(S_+ - S_-) \\ &= \frac{1}{4}((231 \cdot 0,002976) - (105 \cdot 0,002976)) \\ &= \frac{1}{4}(0,6875 - 0,3125) \\ &= \frac{1}{4}(0,375) \\ &= 0,09375 \end{aligned}$$

Kemudian lakukan perhitungan mencari bobot baru untuk data yang terklasifikasi dengan benar dan data salah klasifikasi menggunakan persamaan 3 dan 4.

$$\begin{aligned} w_k &= w_{k-1} \times \frac{S_+ - \text{midpoint}}{S_+} \\ &= 0,002976 \times \frac{0,6875 - 0,1785}{0,6875} \\ &= 0,002976 \times \frac{0,59375}{0,6875} \\ &= 0,002976 \times 0,8636 \\ &= 0,0257 \\ w_k &= w_{k-1} + \frac{\text{midpoint}}{N_-} \\ &= 0,002976 + \frac{0,09375}{105} \\ &= 0,002976 + 0,0893 \\ &= 0,003869 \end{aligned}$$

Dari perhitungan diatas, dapat disimpulkan *boosting* belum dapat dihentikan karena tidak ada salah satu nilai yang memenuhi syarat. Sehingga perlu dilakukan pengacakan data *training* untuk membentuk model pohon keputusan yang baru.

Setelah dilakukan 4 kali iterasi *boosting* dalam pengacakan data *training* secara *random*, menghasilkan model yang mampu mengklasifikasi 313 data dengan benar dan mengalami salah klasifikasi sebanyak 23 data. Dari data *training* tersebut menghasilkan model pohon keputusan sebagai berikut.



Gbr 3. Pohon Keputusan *Boosting*

Dari model tersebut dilakukan perhitungan bobot untuk menentukan apakah perlu dilakukan pengacakan data *training* lagi atau iterasi berhenti di karenakan salah satu syarat pemberhentian *boosting* sudah terpenuhi. Berikut ini perhitungan bobot awal, *midpoint*, bobot baru benar klasifikasi dan salah klasifikasi untuk model terakhir yang terbentuk.

$$w_0 = \frac{1}{N} = \frac{1}{336} = 0,002976$$

$$\begin{aligned} \text{midpoint} &= \frac{1}{4}(S_+ - S_-) \\ &= \frac{1}{4}((313 \cdot 0,002976) - (23 \cdot 0,002976)) \\ &= \frac{1}{4}(0,9315 - 0,0684) \\ &= \frac{1}{4}(0,8631) \\ &= 0,21577 \end{aligned}$$

$$\begin{aligned} w_k &= w_{k-1} \times \frac{S_+ - \text{midpoint}}{S_+} \\ &= 0,002976 \times \frac{0,9315 - 0,21577}{0,9315} \\ &= 0,002976 \times \frac{0,7157}{0,9315} \\ &= 0,002976 \times 0,7683 \\ &= 0,00229 \end{aligned}$$

$$\begin{aligned} w_k &= w_{k-1} + \frac{\text{midpoint}}{N_-} \\ &= 0,002976 + \frac{0,21577}{23} \\ &= 0,002976 + 0,0093811 \\ &= 0,123571 \end{aligned}$$

Dari perhitungan tersebut, dapat disimpulkan iterasi *boosting* dapat dihentikan karena nilai *S* pada perhitungan *midpoint* bernilai kurang dari 0,1 sebesar 0,0684. Oleh karena itu, proses pengacakan data *training* dapat dihentikan. Dan model pohon keputusan ini dijadikan sebagai acuan dalam pengambilan keputusan pada sistem yang dibangun.

C. Pengujian Sistem

Pengujian yang pertama dilakukan dengan melibatkan 144 data *testing* dan diujikan pada model pohon keputusan awal sebelum *boosting*, menghasilkan data sejumlah 91 diklasifikasikan dengan benar dan 53 data masuk ke dalam kategori salah klasifikasi yang disajikan dalam lampiran 4. Dari 91 data yang diklasifikasikan dengan benar terbagi kedalam 3 kelas *True Positive* (TP), diantaranya 33 data TP kelas H, 10 data TP kelas M, dan 48 data kelas L. Untuk lebih jelasnya, dapat dilihat pada tabel di bawah ini.

TABEL 2
CONFUSION MATRIX PENGUJIAN AWAL

		Prediksi		
		H	M	L
Aktual	H	33	15	0
	M	16	10	10
	L	3	9	48

Dari tabel di atas, dilakukan perhitungan *confusion matrix* untuk mencari nilai *recall*, *precision*, dan *accuracy* menggunakan persamaan (5), (6), dan (7) sebagai berikut.

$$\begin{aligned} 1. \text{ Recall H} &= \frac{33}{33+16+3} * 100\% = 63,46\% \\ \text{ Recall M} &= \frac{10}{15+10+9} * 100\% = 29,41\% \\ \text{ Recall L} &= \frac{48}{0+10+48} * 100\% = 82,76\% \\ \text{ All Recall} &= \frac{63,46+29,41+82,76}{3} \% = 58,54\% \\ 2. \text{ Precision H} &= \frac{33}{33+15+0} * 100\% = 68,75\% \\ \text{ Precision M} &= \frac{10}{16+10+10} * 100\% = 27,78\% \\ \text{ Precision L} &= \frac{48}{3+9+48} * 100\% = 80\% \\ \text{ All Precision} &= \frac{68,75+27,78+80}{3} \% = 58,84\% \\ 3. \text{ Accuracy} &= \frac{33+10+48}{144} * 100\% = 63,19\% \end{aligned}$$

Setelah selesai dilakukan pengujian awal sistem, dilanjutkan dengan pengujian sistem yang kedua pada model pohon keputusan setelah diterapkan *boosting* menggunakan 144 data *testing* yang terbagi kedalam 3 kelas sebanyak 48 data dalam kelas H, 36 data kelas M, dan sisanya sebanyak 60 data merupakan kelas L. Dari keseluruhan data *testing* tersebut, sebanyak 48 data pada kelas H sistem mampu memprediksi 45 data kelas H, 2 data pada kelas L, dan 1 data kelas M, untuk 36 kelas M sistem mengklasifikasikan 3 data kelas H, 22 data kelas M, dan 11 data pada kelas L, serta pada 60 data dengan kelas L sistem menghasilkan 4 data kelas H, 2 data kelas M dan 54 data terklasifikasi dengan benar. Selanjutnya dari hasil pengujian tersebut dimasukkan dalam tabel *confusion matrix* seperti pada tabel 3 di bawah ini.

TABEL 3
CONFUSION MATRIX PENGUJIAN BOOSTING

		Prediksi		
		H	M	L
Aktual	H	45	2	1
	M	3	22	11
	L	4	2	54

Kemudian berdasarkan tabel di atas, dilakukan perhitungan untuk mendapatkan nilai *precision*, *recall*, dan *accuracy* dari model yang diujikan menggunakan persamaan (5), (6), dan (7) sebagai berikut.

$$\begin{aligned} 1. \text{ Recall H} &= \frac{45}{45+3+4} * 100\% = 86,54\% \\ \text{ Recall M} &= \frac{22}{2+22+2} * 100\% = 84,62\% \\ \text{ Recall L} &= \frac{54}{1+11+54} * 100\% = 81,82\% \\ \text{ All Recall} &= \frac{86,54+84,62+81,82}{3} \% = 84,32\% \\ 2. \text{ Precision H} &= \frac{45}{45+2+1} * 100\% = 93,75\% \\ \text{ Precision M} &= \frac{22}{3+22+11} * 100\% = 61,11\% \end{aligned}$$

$$\text{Precision L} = \frac{54}{4+2+54} * 100\% = 90\%$$

$$\text{All Precision} = \frac{93,75+61,11+90}{3} \% = 81,62\%$$

$$3. \text{ Accuracy} = \frac{121}{144} * 100\% = 84,03\%$$

Dari hasil perhitungan menggunakan *confusion matrix* di atas, dapat disimpulkan sistem mampu memprediksi data lebih optimal setelah diterapkan *boosting* pada pembentukan model pohon ditandai dengan peningkatan akurasi hasil dari pengujian *confusion matrix* kedua model.

IV. KESIMPULAN

Dari penelitian yang telah dilakukan, dapat disimpulkan fase *boosting* dapat meningkatkan akurasi model pohon keputusan. Dari pengujian yang telah dilakukan, model pohon keputusan yang terbentuk dari data *training* yang menerapkan fase *boosting* mengalami peningkatan akurasi. Namun tidak selalu proses *boosting* menghasilkan kondisi yang lebih akurat, hal ini disebabkan dengan adanya proses pembentukan data *training* baru secara *random*, segala kemungkinan dapat terjadi, dataset yang dihasilkan dapat lebih optimal, memperbanyak data benar klasifikasi dan meminimalisir data salah klasifikasi, atau sebaliknya.

UCAPAN TERIMA KASIH

Penulis menyampaikan ucapan terima kasih kepada seluruh pihak yang berkontribusi dalam memberikan dukungan selama ini sehingga penulis dapat menyelesaikan penelitian dan jurnal ini tepat waktu.

REFERENSI

- [1] Nasution, M. H., & Sofyan, M. I. (2020, Januari). Penerapan Metode Decision Tree Untuk Memprediksi Prestasi Siswa Kelas XII Dilihat dari Nilai Akhir Semester di SMK Negeri 1 Selong Tahun Pelajaran 2017/2018. *Infotek : Jurnal Informatika dan Teknologi, Vol. 3 No. 1*, pp. 58-65.
- [2] Yusuf, B., Qalbi, M., Basrul, Dwitawati, I., Malahayati, & Ellyadi, M. (2020). Implementasi Algoritma Naive Bayes dan Random Forest dalam Memprediksi Prestasi Akademik Mahasiswa Universitas Islam Negeri Ar-Raniry Banda Aceh. *Jurnal Pendidikan Teknologi Informasi, Volume 4, Nomor 1*, 50-58.
- [3] Nevianing, B. C., Mandyartha, E. P., & Nugroho, B. (2021, Maret). Diagnosis ADHD Berdasarkan DSM-IV Menggunakan Algoritma Naive Bayes. *Jurnal Informatika dan Sistem Informasi (JIFoSI), Vol. 2, No. 1*, 1-8.
- [4] Suntoro, J. (2019). *Data Mining : Algoritma dan Implementasi dengan Pemrograman PHP*. Jakarta: PT. Elex Media Komputindo.
- [5] Kastawan, P. W., Wiharta, D. M., & Sudarma, I. M. (2018, September-Desember). Implementasi Algoritma C5.0 pada Penilaian Kinerja Pegawai Negeri Sipil. *Majalah Ilmiah Teknologi Elektro, Vol. 17 No. 3*.
- [6] Yusuf, Y. (2007). Perbandingan Performansi Algoritma Decision Tree C5.0, CART, dan CHAID: Kasus Prediksi Status Resiko Kredit di Bank X. *Seminar Nasional Aplikasi Teknologi Informasi 2007 (SNATI 2007)*.
- [7] Meo, M. O. (2017, Juli). Penerapan Algoritma Pohon Keputusan C5.0 untuk Klasifikasi Lahan Gambut yang Terbakar di Kabupaten Ogan Komering Ilir. *Jurnal Teknologi Terpadu, Vol. 3, No. 1*, pp. 17-23.
- [8] Khun, M., & Johnson, K. (2013). *Applied Predictive Modeling. Springer Science+Business Media*.
- [9] Handono, S. F., Anggraeny, F. T., & Rahmat, B. (2021, Maret). Implementasi Convolutional Neural Network (CNN) Untuk Deteksi Retinopati Diabetik. *Jurnal Informatika dan Sistem Informasi (JIFoSI), Vol. 1, No. 1*, 669-678.
- [10] Amrieh, E. A., Hamtini, T., & Aljarah, I. (2016). Mining Educational Data to Predict Student's Academic Performance using Ensemble Methods. *International Journal of Database Theory and Application, Vol.9, No.8*, pp.119-136.