

## **KETIDAKTEPATAN WAKTU KELULUSAN MAHASISWA UNIVERSITAS TERBUKA DENGAN METODE BOOSTING CART**

**Gede Suwardika<sup>1</sup>, I Ketut Putu Suniantara<sup>2</sup>, Ni Putu Nanik Hendayanti<sup>3</sup>**

<sup>1</sup>Prodi Statistika – UPBJJ Denpasar – Universitas Terbuka

<sup>2,3</sup>Program Studi Sistem Informasi- STMIK STIKOM Bali

*e-mail:* isuwardika@ecampus.ut.ac.id<sup>1</sup>, suniantara@stikom-bali.ac.id<sup>2</sup>,

nanik@stikom-bali.ac.id<sup>3</sup>

### **Abstract**

*The classification tree method or better known as Classification and Regression Tree (CART) has capabilities in various data conditions, but CART is less stable in changing learning data which will cause major changes in the results of the classification tree prediction. Predictive accuracy of an unstable classifier can be corrected by a combination method of many single classifiers where the prediction results of each classifier are combined into the final prediction through the majority voting process for classification or average voting for regression cases. Boosting ensemble method is one method that combines many classification trees to improve stability and determine classification predictions. This research purpose to improve the stability and predictive accuracy of CART with boosting. The case used in this study is the classification of inaccuracies in the Open University student graduation. The results of the analysis show that boosting is able to improve the accuracy of the classification of the inaccuracy of student graduation which reaches a classification prediction of 75.94% which previously reached 65.41% in the classification tree.*

**Key words:** *classification, CART, boosting, ensemble*

### **Abstrak**

Metode pohon klasifikasi atau yang lebih dikenal dengan Classification and Regression Tree (CART) memiliki kemampuan dalam berbagai kondisi data, akan tetapi CART pada hasil prediksi pohon klasifikasi akan mengalami perubahan besar bila terjadi perubahan data learning. Akurasi prediksi dari *classifier* yang tidak stabil dapat diperbaiki dengan metode kombinasi banyak *classifier* tunggal dimana hasil prediksi masing-masing *classifier* digabungkan menjadi prediksi akhir melalui proses voting mayoritas untuk klasifikasi atau voting rata-rata untuk kasus regresi. Mengurangi perubahan dan meningkatkan hasil akurasi prediksi klasifikasi CART dengan mengkombinasikan beberapa klasifikasi dapat menggunakan Metode *boosting ensemble*. Penelitian ini bertujuan untuk mengurangi perubahan dan meningkatkan akurasi hasil prediksi CART dengan *boosting*. Studi kasus dalam penelitian ini adalah klasifikasi ketidaktepatan waktu kelulusan mahasiswa Universitas Terbuka. Hasil analisis menunjukkan bahwa *boosting* mampu meningkatkan akurasi klasifikasi ketidaktepatan waktu kelulusan mahasiswa yang mencapai prediksi klasifikasi mencapai 75,94% yang sebelumnya pada pohon klasifikasi mencapai 65,41%.

**Kata kunci:** *klasifikasi, CART, boosting, ensemble*

---

## I. PENDAHULUAN

Metode Statistika *Classification and Regression Trees* (CART) digunakan untuk mengelompokan suatu data yang tersusun secara sistematis. Metode klasifikasi ini bermasalah ketika variabel – variabel yang digunakan memiliki sejumlah ukuran yang berbeda – beda terdiri dari satu atau beberapa kategori. Metode ini merupakan klasifikasi nonparametrik.

Penerapan metode ini telah digunakan dalam penyelesaian masalah di berbagai bidang seperti bidang kesehatan, marketing, sosial, financial, dan lain sebagainya. Metode ini mampu menghadapi berbagai kondisi data. CART memiliki banyak kelebihan diantaranya mampu mengeksplorasi data berdimensi tinggi dengan komputasi yang efisien, dapat digunakan pada kombinasi data kontinu maupun kategorik, serta interpretasi yang mudah dilakukan. Diantara banyak kelebihan CART, metode klasifikasi CART memiliki kekurangan pada kestabilan dan hasil prediksi pohon klasifikasi ketika terjadi perubahan data learning [1].

*Ensemble* merupakan metode kombinasi banyak *classifier* tunggal dimana hasil prediksi masing-masing *classifier* digabungkan menjadi prediksi akhir melalui proses voting mayoritas untuk klasifikasi atau voting rata-rata untuk kasus regresi. Metode *ensemble* tersebut dikembangkan untuk meningkatkan akurasi prediksi dari *classifier* yang tidak stabil [2]. Penelitian sebelumnya menunjukkan bahwa metode *ensemble* seringkali menghasilkan prediksi yang lebih akurat dibandingkan dengan *classifier* tunggal. Boosting merupakan metode *ensemble* yang relatif baru namun telah menjadi populer. Boosting dikembangkan Freund dan Schapire pada tahun 1996. Karakteristik data menentukan kinerja boosting. Hal ini menyebabkan adanya variasi pada tingkat akurasi klasifikasi [3].

Berbagai penelitian mengenai klasifikasi kelulusan yang telah dilakukan mengindikasikan bahwa banyak sekali faktor yang mempengaruhi ketepatan klasifikasi kelulusan mahasiswa. Menurut [4] Faktor-faktor tersebut dapat bersumber dari diri mahasiswa (faktor internal) atau dari luar diri mahasiswa (lingkungan/faktor eksternal).

Menurut [5] dalam penelitiannya yang berjudul “identifikasi faktor-faktor yang ketepatan waktu kelulusan mahasiswa program pasca sarjana IPB” dengan menggunakan analisis regresi logistik biner bahwa faktor yang mempengaruhi ketepatan kelulusan adalah fakultas, jenis kelamin, jalur masuk, IPK, dan beasiswa dengan akurasi klasifikasi 85%.

Penelitian tentang klasifikasi ketidaktepatan waktu lulusan lainnya dilakukan oleh [6] dalam penelitiannya yang berjudul “analisis *classification and regression tree* (CART) pada lama studi mahasiswa STIKOM Bali menyatakan bahwa faktor yang memengaruhi lama studi mahasiswa adalah lama penyusunan skripsi, IPK dan jurusan mahasiswa dengan ketepatan klasifikasi 87%, akan tetapi hasil klasifikasi tidak stabil disaat ada penambahan variabel dan jumlah data. Menurut [7] mendapatkan hasil akurasi klasifikasi CART 94,2% dan hasil akurasi klasifikasi regresi logistik biner sebesar 86,7% pada data profil mahasiswa FMIPA Universitas Brawijaya. Sedangkan menurut [8] dalam penelitiannya mengatakan bahwa terjadi peningkatan akurasi klasifikasi sebesar 4,38% dengan menggunakan *bagging* CART. Penelitian tentang klasifikasi CART pernah dilakukan oleh [9] pada ketidaktepatan waktu kelulusan dengan analisis random forest menyatakan bahwa random forest mampu meningkatkan akurasi klasifikasi ketidaktepatan waktu kelulusan mahasiswa yang mencapai konvergen dengan prediksi klasifikasi mencapai 93.23%.

Penelitian klasifikasi dengan Metode Arcing CART pernah dilakukan oleh [10] yang disimpulkan bahwa karakteristik utama rumah tangga miskin ialah pekerjaan di bidang pertanian, rendahnya pendidikan, serta banyaknya anggota rumah tangga dengan akurasi klasifikasi sebesar 66,4%. Metode *ensemble* Arcing yang dilakukan dapat meningkatkan akurasi klasifikasi sebesar 12,7% menjadi 79,1%. Menurut [11] melakukan klasifikasi kemiskinan di Provinsi Jawa Timur dengan Boosting MARS yang disimpulkan bahwa karakteristik yang menentukan daerah miskin ialah penggunaan jamban, rata-rata lama sekolah, sektor pekerjaan, dan angka melek huruf suatu

daerah. Metode *ensemble* Boosting yang dilakukan dapat meningkatkan akurasi MARS sebesar 89,47% menjadi 97,37%. Klasifikasi dengan metode *ensemble* mampu memberikan akurasi yang lebih baik pada klasifikasi kemiskinan masyarakat di Kabupaten Jombang. Dalam hal hasil prediksi pohon klasifikasi random forest merupakan metode terbaik, sedangkan *bagging* CART merupakan metode terbaik pada kecepatan akurasi klasifikasi untuk mencapai konvergen [12].

Berdasarkan latarbelakang masalah tersebut di atas diterapkan metode *ensemble boosting* pada CART dengan tujuan meningkatkan kestabilan dan akurasi prediktif pohon yang dihasilkan. Hasil klasifikasi yang diperoleh akan dibandingkan dengan metode CART (tanpa boosting) untuk mendapatkan model klasifikasi terbaik.

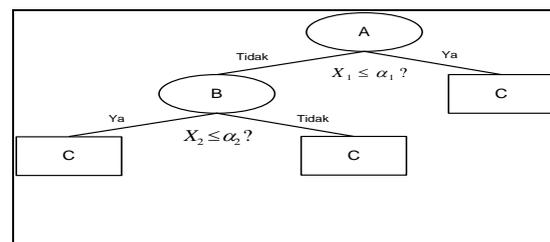
#### a. Classification And Regression Trees (CART)

Regresi berstruktur pohon menggunakan kaidah pohon keputusan (*decision tree*) dibentuk melalui algoritma penyekatan rekursif. Metode ini dikembangkan oleh Morgan dan Sonquist pada awal tahun 1960-an dengan bantuan program AID (*Automatic Interaction Detection*). Kajian mengenai metode pohon dimulai pada tahun 1973, yaitu pada saat Breiman dan Friedman secara terpisah menggunakan metode pohon ini dalam klasifikasi. Perkembangan metode ini ditandai dengan diterbitkannya buku *Classification and Regression Trees* pada tahun 1984 [13] [14] [15].

Kelompok – kelompok pengamatan ditentukan berdasarkan variabel respon yang dipengaruhi oleh variabel penjelas. Penyekatan rekursif pada ruang variabel penjelas  $X$  diperoleh melalui suatu algoritma. Hasil penyekatan ini menjadi struktur dalam regresi pohon [16]. Penyekatan dimulai dengan menyekat variabel penjelas menjadi dua anak gugus (disebut simpul). Selanjutnya setiap simpul ini disekat lagi menjadi dua anak simpul baru. Proses ini diulang sampai diperoleh sekatan yang berdasarkan suatu aturan tertentu tidak dapat disekat lebih lanjut. Melalui proses ini, diperoleh sekatan-sekatan dengan perbedaan yang sangat besar dengan respon yang lebih homogen dalam tiap-tiap sekatan [17].

Hasil representasi suatu struktur pohon dapat dilihat pada Gambar 1 yang menampilkan sebuah struktur pohon sederhana yang mempunyai tiga lapisan simpul [17], yaitu:

1. Lapisan pertama (A) yang merupakan simpul induk atau simpul akar (*root node*).
2. Lapisan kedua (B dan C) yang merupakan simpul anak yang terdiri dari satu simpul internal (*internal node*) (lingkaran) dan satu simpul terminal (kotak).
3. Lapisan ketiga (C) yang merupakan simpul akhir yang tidak bercabang lagi yang terdiri dari dua simpul terminal (kotak).



Gambar 1. Struktur Pohon Regresi

Variabel penjelas digambarkan oleh lapisan A, B, dan C yang terpilih untuk menjadi simpul pembentuk struktur pohon. Nilai tengah antara dua nilai amatan variabel  $x_i$  berturut yaitu  $\alpha_1$  dan  $\alpha_2$ . Misalkan terdapat  $n$  variabel penjelas,  $X_1, X_2, \dots, X_n$  dengan satu variabel respon kontinu. Beberapa komponen penting dalam pembentukan pohon regresi adalah:

1. Pertanyaan yang bersifat dikotomi berupa pertanyaan ya atau tidak yang dibentuk dalam himpunan  $x_i \in A?$  dengan  $x_i$  merupakan suatu amatan contoh dari  $A \subset X$  (ruang variabel penjelas). Sekatan ruang variabel penjelas ditentukan oleh jawaban pertanyaan tersebut. Amatan dengan jawaban “ya” akan masuk ke ruang A sedangkan amatan dengan jawaban “tidak” akan masuk ke ruang  $A^c$  (A komplemen). Sekatan ruang sampel yang terbentuk disebut simpul (*node*).
2. Kriteria kebaikan sekatan (*goodness of split*),  $\phi(s, g)$ , sebagai alat evaluasi bagi baik tidaknya penyekatan yang dilakukan oleh sekatan  $s$  terhadap simpul  $g$ .

3. Aturan penghentian (*stopping rule*) untuk menentukan kapan penyekatan suatu simpul harus dihentikan.
4. Statistik yang digunakan sebagai ringkasan dari tiap-tiap simpul akhir.

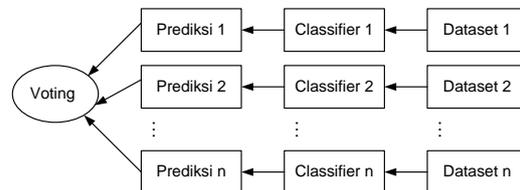
Himpunan pertanyaan dikotomus  $Q$  dapat dilakukan sebagai berikut:

- a. Tiap sekatan tergantung pada nilai satu variabel penjelas.
- b. Untuk variabel kontinu  $X_i$ , Pemilah hanya berasal dari pertanyaan yang dibakukan dalam bentuk "apakah  $X_i \leq c$ ", untuk  $c \in R^1$ . Jadi, jika ruang contohnya berukuran  $n$  dan terdapat sebanyak-banyaknya  $n$  nilai amatan berbeda pada variabel  $X_i$ , maka akan terdapat sebanyak-banyaknya  $n-1$  *split* yang berbeda yang dibentuk oleh gugus pertanyaan {"apakah  $X_i \leq c_i$ ?"}, dengan  $i = 1, 2, \dots, n-1$  dan  $c_i$  adalah nilai tengah antara dua nilai amatan variabel  $X_i$  berurutan dan berbeda.
- c. Untuk variabel penjelas yang berjenis katagorik, pemillihan yang terjadi berasal dari semua kemungkinan pemillihan terbentuk dari dua anak gugus yang saling lepas. Jika  $X_i$  bernilai  $b \in \{b_1, b_2, \dots, b_L\}$ , pertanyaan berbentuk "apakah  $X_i \in B$ ", dimana  $B$  merupakan semua himpunan bagian dari  $\{b_1, b_2, \dots, b_L\}$ . Maka variabel  $X_i$  merupakan variabel berskala nominal dengan taraf  $L$  maka akan terdapat sebanyak  $2^{L-1} - 1$  pertanyaan yang mungkin. Jika variabel penjelas  $X_i$  merupakan variabel yang diskontinu atau ordinal dengan  $L$  nilai yang berbeda, maka akan terdapat sebanyak  $L-1$  pertanyaan yang mungkin.

b. Metode *Esemble Boosting*

Metode *ensemble* dikembangkan dengan harapan untuk meningkatkan akurasi

klasifikasi dari sebuah *classifier* tunggal. Ide dasarnya adalah dengan menggunakan banyak *classifier* dari metode yang sama dan kemudian mengkombinasikannya melalui proses voting untuk memperoleh dugaan klasifikasi akhir (Wezel dan Potharst, 2007). Ilustrasi umum metode *ensemble* ialah sebagai berikut:



Gambar 2. Ilustrasi Metode *Ensemble*

Boosting merupakan salah satu metode *ensemble* yang meliputi banyak algoritma, dimana *AdaBoost* merupakan salah satu yang paling populer. Metode ini pertama kali dikembangkan oleh Freund dan Schapire pada Tahun 1995. Boosting CART secara umum berfokus untuk membuat sebuah deret pohon klasifikasi. Data set yang digunakan pada setiap pohon klasifikasi bergantung pada pohon klasifikasi sebelumnya dan fokus pada data yang salah diprediksi. Data yang salah diprediksi akan diperbaiki terus menerus oleh pohon klasifikasi selanjutnya. Menurut [18] menyatakan bahwa Boosting merupakan salah satu metode *ensemble* yang paling baik.

c. Ukuran Akurasi Klasifikasi

Untuk melihat ketepatan seberapa besar pengelompokan tersebut tepat pada kelompoknya dapat digunakan akurasi klasifikasi. Perhitungan akurasi klasifikasi untuk menghitung nilai 1-APER (*apparent error rate*) akan menghasilkan peluang kesalahan klasifikasi yang kecil. Perhitungan hasil klasifikasi dengan peluang kecil dikatakan sebagai metode klasifikasi terbaik. APER merupakan sejumlah observasi dataset yang salah diklasifikasikan oleh sebuah fungsi klasifikasi. Penghitungan APER dapat diketahui melalui tabel klasifikasi berikut:

Tabel 1. Klasifikasi

		Kelas Sebenarnya	
		$y_1 = -1$	$y_2 = 1$
Kelas Prediksi	$h_1 = -1$	TN <i>true negative</i>	FN <i>false negative</i>
	$h_2 = 1$	FP	TP

*false positive      true positive*

dimana:

TN : jumlah  $h_1$  dari hasil klasifikasi yang tepat disklasifikasi sebagai  $y_1$

FP : jumlah  $h_2$  dari hasil klasifikasi yang salah disklasifikasi sebagai  $y_1$

FN : jumlah  $h_1$  dari hasil klasifikasi yang salah disklasifikasi sebagai  $y_2$

TP : jumlah  $h_2$  dari hasil klasifikasi yang tepat disklasifikasi sebagai  $y_2$

Maka perhitungan akurasi klasifikasi dengan 1-APER sebagai berikut:

$$1 - APER = \frac{TP+TN}{TP+FP+TN+FN} \quad (1)$$

Ukuran stabilitas klasifikasi metode *ensemble* diukur menggunakan standar deviasi  $1 - APER$ . Akurasi klasifikasi kelas positif digambarkan oleh nilai dari *Sensitivity*, sedangkan akurasi klasifikasi pada kelas negative digambarkan oleh *specificity*. Nilai *G-means* menggambarkan ukuran klasifikasi antara nilai *Sensitivity* dan *specificity*. *G-means* baik digunakan untuk mengukur akurasi prediksi metode klasifikasi untuk data yang tidak seimbang.

$$sensitivity = TPR = \frac{TP}{TP+FN} \quad (2)$$

$$specificity = \frac{TN}{TN+FP} \quad (3)$$

$$1 - specificity = FPR = \frac{FP}{TN+FP} \quad (4)$$

$$G - means = \sqrt{sensitivity \times specificity} \quad (5)$$

Metode klasifikasi dapat dikatakan baik jika mampu mengukur nilai *sensitivity* dan *specificity*. *G-means* dapat menggambarkan bagaimana sebuah klasifikasi mampu mengukur *sensitivity* dan *specificity*.

## II. METODE PENELITIAN

Penelitian ini dilakukan di Universitas Terbuka UPPJJ Denpasar dengan waktu penelitian sepuluh bulan. Adapun variabel – variabel yang digunakan yaitu 1). Variabel respon (Y) yang menyatakan status kelulusan mahasiswa tepat waktu dan tidak tepat waktu, 2). Variabel bebasnya yaitu terdiri dari: Jenis Kelamin laki – laki dan perempuan ( $X_1$ ) Jurusan/prodi yang terdiri dari 6 jurusan ( $X_2$ ), IPK ( $X_3$ ), Pendidikan Orang Tua ( $X_4$ ), pekerjaan orang tua (bapak dan ibu) ( $X_5$ ),

penghasilan orang tua ( $X_6$ ) dan status mahasiswa bekerja dan tidak berkerja ( $X_7$ ). Adapun tahapan – tahapan penelitian ini sebagai berikut:

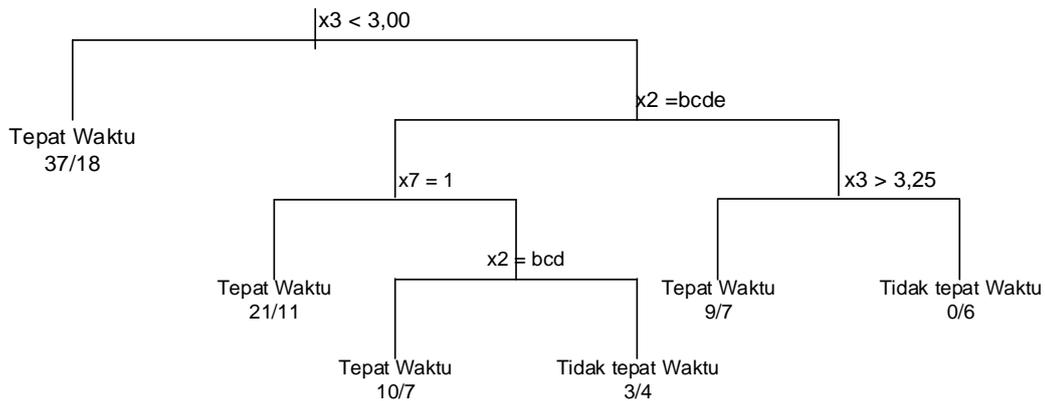
1. Analisis CART, dengan tahapan sebagai berikut:
  - a. Menentukan semua kemungkinan penyekatan pada tiap variabel penjelas.
  - b. Menghitung kehomogenan simpul pada semua variabel penjelas.
  - c. Anak simpul kiri dan kanan dibentuk oleh variabel penjelas berdasarkan data yang dibentuk oleh simpul induk yang terbentuk terlebih dahulu.
  - d. Pembentukan pohon klasifikasi berhenti sampai aturan pembentukan pohon terpenuhi.
  - e. Pemangkasan pohon dilakukan untuk mendapatkan pohon akhir yang lebih sederhana.
  - f. Setelah diperoleh pengelompokan, kemudian melakukan interpretasi.
2. Analisis klasifikasi Boosting CART dengan tahapan sebagai berikut:
  - a. Memberikan bobot pada dataset  $D$
  - b. Membentuk moel CART dari dataset yang telah diboboti
  - c. Mengamati prediksi pengelompokkan untuk tiap sampel data
  - d. Mengulangi langkah *a-c* untuk 2-20 kali replikasi dimana bobot akan semakin besar untuk data yang salah dikelompokkan dan semakin kecil untuk data yang benar dikelompokkan oleh iterasi sebelumnya
  - e. Melakukan voting mayoritas terhadap dugaan pengelompokan tiap data sampel dari 2-20 replikasi yang telah dilakukan
  - f. Menghitung  $1 - APER$  klasifikasi *Boosting CART*, *sensitivity*, *specificity*, dan *G-means*.

## III. HASIL dan PEMBAHASAN

### 1. Classification And Regression Trees

Metode *Classification And Regression Trees* (CART) menggunakan algoritma pohon keputusan (*decision tree*). Hasil analisis CART berupa pohon klasifikasi pada kasus ketepatan

waktu lulusan mahasiswa Universitas Terbuka ditunjukkan pada Gambar 3.



Gambar 3. Pohon Klasifikasi CART

Berdasarkan pohon klasifikasi tersebut diketahui bahwa variabel yang digunakan sebagai pemilah pohon klasifikasi CART dan paling menentukan waktu kelulusan yaitu IPK ( $x_3$ ), program studi ( $x_2$ ), Status mahasiswa ( $x_7$ ). Klasifikasi ini terbagi dalam dua kelas, yaitu lulus tepat waktu dan tidak tepat waktu.

a. Kelulusan Tepat Waktu, lulus tepat waktu menjadi 4 simpul dengan karakteristik sebagai berikut:

1. Kelulusan tepat waktu dengan IPK kurang dari 3.
2. Kelulusan tepat waktu dengan IPK lebih dari 3; program studi selain ilmu administrasi Negara; dengan status mahasiswa bekerja
3. Kelulusan tepat waktu dengan IPK lebih dari 3; program studi selain ilmu administrasi Negara dan ilmu komunikasi; dengan status mahasiswa bekerja
4. Kelulusan tepat waktu dengan IPK lebih dari 3,25; program studi selain akuntansi; dengan status mahasiswa tidak bekerja.

b. Kelulusan Tidak Tepat Waktu, lulus tidak tepat waktu menjadi 2 simpul dengan karakteristik sebagai berikut:

1. Kelulusan tidak tepat waktu dengan IPK lebih dari 3; status mahasiswa bekerja pada program studi manajemen.
2. Kelulusan tidak tepat waktu dengan IPK lebih dari 3,25; program studi selain akuntansi; dengan status mahasiswa tidak bekerja.

Berdasarkan pohon klasifikasi CART didapatkan hasil klasifikasi pada Tabel 2. Tabel 2 klasifikasi CART menyatakan bahwa terdapat sebanyak 77 data sampel waktu kelulusan diklasifikasikan tepat waktu sedangkan 3 lainnya salah diklasifikasikan. Pada data sampel tidak tepat waktu, terdapat 10 data sampel tepat diklasifikasikan tidak tepat waktu namun data sampel yang salah diklasifikasikan lebih banyak yaitu 43. Klasifikasi digunakan untuk mengukur akurasi klasifikasi pada pohon klasifikasi CART dimana dapat dihitung nilai 1-APER, *sensitivity*, *specificity*, dan *G-means*.

Tabel 2. Klasifikasi CART

		Kelas Sebenarnya	
		Tepat waktu	tidak tepat waktu
Kelas Prediksi	tepat waktu	77	43
	tidak tepat waktu	3	10

Hasil perhitungan menunjukkan bahwa 1-APER pohon klasifikasi CART bernilai 0,6541 atau dengan kata lain, data sampel yang tepat diklasifikasikan secara keseluruhan sebanyak 65,41%. *Sensitivity* pohon klasifikasi CART bernilai 0,1887 yang artinya hanya 18,87% data waktu kelulusan tepat diklasifikasikan tidak tepat waktu. Adapun *specificity* bernilai 0,9625 yang artinya sebanyak 96,25% waktu lulusan tepat diklasifikasikan tepat waktu. Hal tersebut menunjukkan bahwa CART hanya bagus dalam mengklasifikasikan tepat waktu kelulusan mahasiswa. Sedangkan akurasi klasifikasi untuk waktu kelulusan tidak tepat waktu sangat rendah. Berikut merupakan perhitungan nilai 1-APER, *sensitivity*, *specificity*, dan G-means hasil klasifikasi dengan CART.

$$1 - APER = \frac{TP + TN}{TP + FP + TN + FN} = \frac{10 + 77}{133} = 0,6541$$

$$sensitivity = \frac{TP}{TP + FN} = \frac{10}{53} = 0,1887$$

$$specificity = \frac{TN}{TN + FP} = \frac{77}{80} = 0,9625$$

$$G - measn = \sqrt{sensitivity \times specificity} = \sqrt{0,1887 \times 0,9626} = 0,4261$$

Pengujian signifikansi prediksi pohon klasifikasi CART digunakan uji Press's Q, dengan jumlah pengamatan total  $N = 133$ , jumlah pengamatan yang benar diklasifikasikan  $n = 80$ , dan jumlah kelas  $K = 2$ , maka perhitungan statistik Press's Q bernilai 12,6391. Adapun nilai kritis pada tingkat signifikansi 0,01 ialah 6,63.

$$Press's Q = \frac{[N - (nK)]^2}{N(K - 1)} = \frac{[133 - (87.2)]^2}{133(2 - 1)} = 12,6391$$

Berdasarkan statistik *Press's Q*, didapatkan nilainya lebih besar dari nilai kritis. Sehingga dapat disimpulkan bahwa prediksi pohon klasifikasi CART signifikan.

## 2. Klasifikasi dengan Boosting CART

Hasil klasifikasi Boosting CART pohon klasifikasi yang dilakukan pada data ketepatan waktu kelulusan mahasiswa Universitas Terbuka, hampir sama dengan menggunakan bagging CART dimana terdapat 3 variabel yang paling sering digunakan sebagai pemilah pohon dan paling menentukan ketepatan kelulusan secara berurutan adalah Indeks Prestasi Kumulati (x3), Program Studi (x2), dan status mahasiswa (x7). Sebanyak 68 data sampel ketepatan waktu kelulusan diklasifikasikan tepat waktu sedangkan 11 lainnya salah diklasifikasikan sebagai ketidaktepatan waktu kelulusan. Pada data sampel tidak tepat waktu, hanya 33 data sampel tepat diklasifikasikan tidak tepat waktu. Sedangkan 21 data sampel salah diklasifikasikan sebagai kelulusan tepat waktu.

Perhitungan nilai 1-APER, *sensitivity*, *specificity*, dan G-means hasil klasifikasi dengan Boosting CART sebagai berikut:

$$1 - APER = \frac{TP + TN}{TP + FP + TN + FN} = \frac{68 + 33}{133} = 0,7594$$

$$sensitivity = \frac{TP}{TP + FN} = \frac{33}{33 + 21} = 0,6111$$

$$specificity = \frac{TN}{TN + FP} = \frac{68}{68 + 11} = 0,8608$$

Tabel 3. Klasifikasi Boosting CART

		Kelas Sebenarnya	
		Tepat waktu	tidak tepat waktu
Kelas Prediksi	tepat waktu	68	21
	tidak tepat waktu	11	33

Hasil perhitungan menunjukkan bahwa 1-APER pohon klasifikasi Boosting CART bernilai 0,7594 atau dengan kata lain, data sampel yang tepat diklasifikasikan secara keseluruhan sebanyak 75,94%. *Sensitivity* pohon klasifikasi Boosting CART bernilai 0,6111 yang artinya hanya 61,11% data waktu kelulusan tepat diklasifikasikan tidak tepat waktu. Adapun *specificity* bernilai 0,8608 yang artinya sebanyak 86,06% waktu lulus diklasifikasikan tepat waktu. Nilai ketiganya tidak jauh berbeda atau relatif berimbang yang artinya Boosting CART bagus dalam mengklasifikasikan baik untuk kelas lulus tepat waktu maupun lulus tidak tepat waktu.

$$G - means = \sqrt{sensitivity \times specificity}$$

$$= \sqrt{0,7594 \times 0,8607}$$

$$= 0,7253$$

Mengukur kecepatan akurasi prediksi untuk mencapai konvergen pada Metode *Ensemble* dilihat dengan menggunakan standar deviasi 1-APER Boosting CART yang diperoleh sebesar 0,0291. Pengujian signifikansi prediksi klasifikasi Bagging CART digunakan uji Press's Q. Diketahui bahwa jumlah pengamatan total  $N = 133$ , jumlah pengamatan yang benar diklasifikasikan  $n = 101$ , dan jumlah kelas  $K = 2$ , maka perhitungan statistik Press's Q

bernilai 35,7970. Adapun nilai kritis pada tingkat signifikansi 0,01 ialah 6,63.

$$Press's Q = \frac{[N - (nK)]^2}{N(K - 1)} = \frac{[133 - (87.2)]^2}{133(2 - 1)}$$

$$= 12,6391$$

Berdasarkan statistik Press's Q, didapatkan nilai yang lebih besar dari nilai kritis. Karena statistik Press's Q melebihi nilai kritis, maka dapat disimpulkan bahwa prediksi Boosting CART signifikan secara statistik.

### 3. Perbandingan Klasifikasi

Akurasi prediksi secara umum diukur dari nilai 1-APER dimana semakin besar 1-APER maka semakin baik akurasi prediksinya. Semakin besar *sensitivity* maka semakin baik akurasi prediksi klasifikasi tidak tepat waktu sedangkan semakin besar *specificity* maka semakin baik akurasi prediksi klasifikasi tepat waktu. Kemampuan dalam prediksi data diukur dari keseimbangan *sensitivity* dan *specificity* yang dirangkum dalam nilai *G-means* dimana semakin besar nilainya maka semakin baik. Sedangkan kecepatan akurasi prediksi untuk mencapai konvergen diukur dari standar deviasi 1-APER dimana semakin kecil semakin cepat atau semakin sedikit pohon yang dibutuhkan untuk mencapai akurasi yang konvergen.

Tabel 4. Ukuran Kinerja Metode Klasifikasi

Metode Klasifikasi	1 - APER	Sensitivity	Specificity	G-Means
CART	0,6541*	0,1887	0,9625	0,4261
Boosting CART	0,7594*	0,6111	0,8608	0,7253

Keterangan: \*Statistik Press's Q signifikan

Berdasarkan Tabel 4 dapat diketahui bahwa 1-APER pohon klasifikasi CART bernilai 0,6541 yang artinya sebanyak 65,41% data sampel tepat diklasifikasikan. Sedangkan *boosting* CART menghasilkan 1-APER bernilai 75,94%. Melihat nilai 1-APER dari kedua metode dapat dikatakan metode *boosting* CART mampu meningkatkan akurasi klasifikasi CART secara umum.

Nilai *Sensitivity* pohon klasifikasi CART dan *boosting* CART bernilai 0,1887 dan 0,6111 menunjukkan artinya hanya 18,87% data ketepatan waktu lulus tepat diklasifikasikan tidak tepat waktu pada

metode CART sedangkan pada *boosting* CART 61,11%. Adapun *specificity* bernilai 0,9625 yang artinya 96,25% ketepatan waktu lulus tepat diklasifikasikan tepat waktu pada klasifikasi CART. Sedangkan pada *boosting* CART 0,8608 atau 86,08%. Angka tersebut menunjukkan bahwa akurasi pohon klasifikasi CART belum cukup bagus dalam memprediksi kelulusan tidak tepat waktu. Adapun akurasi metode *boosting* CART lebih baik dalam memprediksi ketepatan waktu lulus tidak tepat waktu, hal ini dapat dilihat nilai *sensitivity* yang lebih besar.

Keseimbangan akurasi prediksi untuk tiap kelas dapat diukur dari nilai *G-means*. Aspek ini perlu diperhitungkan karena metode klasifikasi cenderung baik dalam memprediksi kelas dengan data sampel yang lebih banyak namun buruk dalam memprediksi kelas dengan data sampel sedikit. Pohon klasifikasi CART memiliki *G-means* bernilai 0,4261. Sedangkan *boosting CART* menghasilkan pohon klasifikasi yang memiliki nilai *G-means* (0,7253) lebih besar dibanding CART yang artinya akurasi prediksi metode *boosting CART* cenderung lebih seimbang dalam melakukan prediksi di tiap kelas.

#### IV. KESIMPULAN dan SARAN

##### 1. Kesimpulan

Berdasarkan pembahasan tersebut diatas, maka kesimpulan yang diperoleh, yaitu:

- a. Penerapan klasifikasi CART pada klasifikasi ketidaktepatan waktu lulusan mahasiswa Universitas Terbuka menunjukkan bahwa waktu ketepatan kelulusan dipengaruhi oleh IPK, status mahasiswa, program studi dan penghasilan keluarga sebagai variabel pemilah pohon klasifikasi CART.
- b. *Boosting CART* menghasilkan akurasi klasifikasi yang lebih baik dibandingkan dengan pohon klasifikasi CART.

##### 2. Saran

Saran yang dapat diberikan dalam hasil penelitian ini adalah menambahkan variabel penjelas lainnya dan membandingkan dengan metode *ensemble* lainnya.

#### V. UCAPAN TERIMA KASIH

Pada kesempatan ini, penulis mengucapkan terima kasih yang sebesar – besarnya kepada Universitas Terbuka telah memberi dukungan finansial terhadap penelitian ini, sehingga penulis dapat menyelesaikan penelitian ini dengan baik dan tepat waktu.

#### VI. DAFTAR PUSTAKA / REFERENSI

- [1] R. Timofeev, "Classification and Regression Trees (CART) Theory and Applications," Humboldt University, Berlin, 2004.
- [2] M. Wezel and R. Potharst, "Improved Customer Choice Predictions using Ensemble Methods," *Eur. J. Oper. Res.*, vol. 181, no. 1, pp. 436–452, 2007.
- [3] D. Opitz and R. Maclin, "Popular Ensemble Methods: An Empirical Study," *J. Artif. Intell. Res.*, vol. 11, pp. 169–198, 1999.
- [4] S. Suryabrata, *Psikologi Pendidikan*. Jakarta: Raja Grafindo Persada, 2008.
- [5] N. Nugenita, "Identifikasi Faktor-faktor Yang Mempengaruhi Ketepatan Waktu Kelulusan Mahasiswa Program Sarjana IPB," IPB Bogor, 2015.
- [6] I. K. P. Suniantara, "Penerapan Metode Regresi Berstruktur Pohon Dalam Memprediksi Berat Badan Bayi Lahir, Studi Kasus: Rumah Sakit Umum Daerah Wangaya," Jurusan Matematika, Universitas Udayana, 2008.
- [7] A. Margasari, "Penerapan metode CART (Classification and Regression Trees) dan analisis regresi logistik Biner pada klasifikasi profil mahasiswa FMIPA Universitas Brawijaya," *Studen J.*, vol. 2, no. 4, 2014.
- [8] I. K. P. Suniantara and M. Rusli, "Ketepatan Klasifikasi Bagging CART Pada Klasifikasi Ketidaktepatan Waktu Kelulusan Mahasiswa STIKOM Bali," in *Konferensi Nasional Sistem dan Informatika*, 2017.
- [9] G. Suwardika and I. K. P. Suniantara, "Analisis Random Forest Pada Klasifikasi Cart Ketidaktepatan Waktu Kelulusan Mahasiswa Universitas Terbuka," in *Seminar Nasional Tahunan Matematika Sains dan Teknologi*, 2018.
- [10] E. Q. Aeni, "Pendekatan CART Arcing untuk Klasifikasi Kesejahteraan Rumah Tangga di Propinsi Jawa Tengah," Institut Teknologi Sepuluh Nopember, Surabaya, 2009.
- [11] E. . Permatasari, "Pendekatan Boosting Multivariate Adaptive Regression Spline (MARS) untuk Klasifikasi Kemiskinan di Propinsi Jawa Timur," Institut Teknologi Sepuluh Nopember Surabaya, 2013.
- [12] M. J. Muttaqin, B. W. Otok, and S. P. Rahayu, "Metode Ensemble pada

- CART untuk Perbaikan Klasifikasi Kemiskinan di Kabupaten Jombang,” Institut Teknologi Sepuluh Nopember Surabaya, 2013.
- [13] L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone, “Classification and Regression Trees,” New York: Chapman and Hall.
- [14] J. Morgan, “Classification and Regression Tree Analysis,” Boston University School of Public Health, 2104.
- [15] W. Y. Loh, “Classification and Regression Tree,” *WIREs Data Min. Knowl. Discov.*, vol. 1, no. 1, 2011.
- [16] W. Y. Loh, “Classification and Regression Tree Methods,” in *In Encyclopedia of Statistics in Quality and Reliability*, Wiley, 2008, pp. 315–323.
- [17] P. Melillo, L. Pecchia, P. A. Bath, and M. Bracale, “The use of Classification and Regression Tree to predict 15-year survival in community-dwelling older people,” *Int. J. Heal. Inf. Manag. Res.*, vol. 1, no. 1, pp. 48–54, 2013.
- [18] D. . Cao, Q. . Xu, Y. . Liang, L. . Zhang, and H. . Li, “The Boosting: A New Idea of Building Models,” *Chemom. Intell. Lab. Syst.*, vol. 100, pp. 1–11, 2010.