

Ketepatan Klasifikasi *Bagging* CART Pada Klasifikasi Ketidaktepatan Waktu Kelulusan Mahasiswa STIKOM Bali

I Ketut Putu Suniantara¹⁾, Muhammad Rusli²⁾

Program Studi Sistem Informasi
STMIK STIKOM Bali

Jl. Raya Puputan No. 68 Renon – Denpasar – Bali. Tlpn (0361) 244445
e-mail: [1suniantara@stikom-bali.ac.id](mailto:suniantara@stikom-bali.ac.id), [2rusli@stikom-bali.ac.id](mailto:rusli@stikom-bali.ac.id)

Abstrak

CART (Classification and Regression Tree) merupakan salah satu metode klasifikasi yang populer digunakan di berbagai bidang. Metode *CART* mampu menghadapi berbagai kondisi data, akan tetapi *CART* mempunyai kelemahan pohon klasifikasi yaitu kurang stabil pada perubahan data *learning* yang akan menyebabkan perubahan besar pada hasil prediksi pohon klasifikasi. Untuk memperbaiki prediksi dari pohon klasifikasi *CART* dari kelemahan tersebut dikembangkan metode *Bagging* (*bootstrap aggregating*). Penelitian ini bertujuan untuk meningkatkan kestabilan dan akurasi prediktif. Sedangkan kasus yang digunakan dalam penelitian ini adalah klasifikasi ketidaktepatan waktu kelulusan mahasiswa STIKOM Bali. Hasil klasifikasi dengan *bagging* *CART* mampu meningkatkan akurasi klasifikasi sebesar 4,38%.

Kata kunci: *CART*, *Bagging*, klasifikasi, kelulusan

1. Pendahuluan

Pengklasifikasian merupakan salah satu metode statistika dalam pengelompokan suatu data yang disusun secara sistematis. Masalah klasifikasi muncul ketika terdapat sejumlah ukuran yang terdiri dari satu atau beberapa kategori yang tidak dapat diidentifikasi secara langsung tetapi harus menggunakan suatu ukuran. Salah satunya metode klasifikasi adalah metode *CART (Classification And Regression Trees)*.

CART merupakan salah satu metode klasifikasi nonparametrik yang populer digunakan. Metode ini telah digunakan dalam penyelesaian masalah diberbagai bidang seperti bidang kesehatan, marketing, sosial, finansial, dan lain sebagainya. Metode ini mampu menghadapi berbagai kondisi data. *CART* memiliki banyak kelebihan diantaranya mampu mengeksplorasi data berdimensi tinggi dengan komputerisasi yang efisien dan dapat digunakan pada kombinasi data kontinu maupun kategorik, serta interpretasi yang mudah dilakukan [1] dan [2]. Diantara banyak kelebihan tersebut, *CART* juga memiliki kelemahan yaitu kurang stabil pada perubahan data *learning* yang akan menyebabkan perubahan besar pada hasil prediksi pohon klasifikasi.

Untuk mengatasi kelemahan tersebut dikembangkan suatu metode *ensemble* yaitu *Bagging* (*bootstrap aggregating*). *Bagging* (*bootstrap aggregating*) dikembangkan untuk meningkatkan akurasi prediksi dari *classifier* yang tidak stabil [3]. *Bagging* (*bootstrap aggregating*) merupakan metode *ensemble* yang relatif baru namun telah menjadi populer. *Bagging* dikembangkan oleh Breiman pada Tahun 90-an. *Bagging* bertujuan untuk mereduksi varians prediktor [4]. Menurut [5] bahwa *Bagging* *CART* mampu meningkatkan ketepatan hasil klasifikasi dalam kecepatan akurasi klasifikasi untuk mencapai konvergen.

Beberapa penelitian tentang ketepatan waktu lulusan mahasiswa dilakukan oleh [6], [7] dan [8] menyatakan bahwa status kelulusan tepat waktu dipengaruhi oleh lama menyusul skripsi, IPK dan Jurusan dengan rata – rata akurasi klasifikasi di atas 90%, akan tetapi dalam akurasi prediksi pada pohon tidak stabil ketika terjadi perubahan pada data *learning*.

Penelitian ini bertujuan untuk meningkatkan kestabilan dan akurasi prediktif *CART* dengan teknik *Bootstrap Aggregating* (*Bagging*). Sedangkan kasus yang digunakan dalam penelitian ini adalah klasifikasi ketidaktepatan waktu kelulusan mahasiswa STIKOM Bali. Hasil yang diharapkan dalam penelitian ini adalah ketepatan klasifikasi dan kecepatan akurasi klasifikasi pada metode tersebut.

2. Metode Penelitian

Penelitian ini dilakukan di STIKOM – Bali dengan waktu penelitian lima bulan, dimana data yang digunakan dalam penelitian ini data lulusan mahasiswa STIKOM BALI periode 2016 – 2017 sebanyak 890 sampel mahasiswa dengan rincian sebanyak 405 sampel mahasiswa kelulusan tepat waktu dan sebanyak 485 sampel mahasiswa kelulusannya tidak tepat waktu. Adapun variabel bebas dan variabel respon diperlihatkan pada Tabel 1.

Tabel 1. Variabel Penelitian

No	Variabel	Jenis	Kategori
1	Status kelulusan mahasiswa	Diskrit	tepat waktu dan tidak tepat waktu
2	Jenis kelamin mahasiswa	Diskrit	Laki-laki dan Perempuan
3	Jurusan/Prodi mahasiswa	Diskrit	SI dan SK
4	Lama skripsi mahasiswa	Kontinu	
5	IPK	Kontinu	
6	IPS semester 6	Kontinu	
7	Nilai ujian masuk	Kontinu	

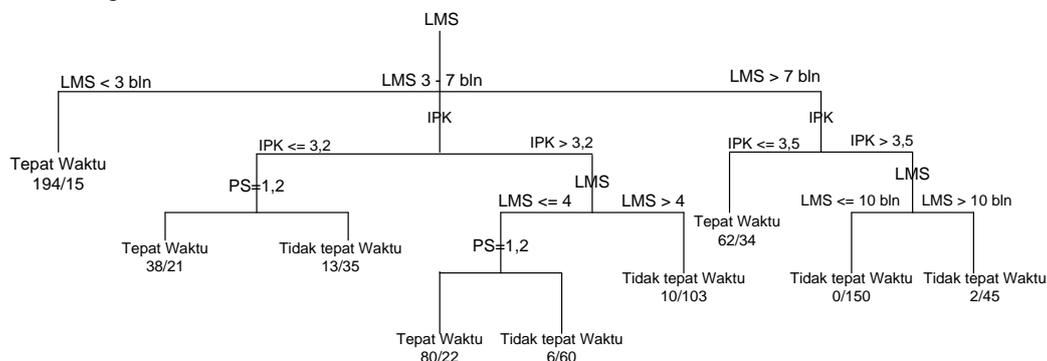
Adapun langkah-langkah untuk mencapai tujuan penelitian tersebut sebagai berikut:

- Bagi data penelitian menjadi dua bagian yaitu data *learning* (L) dan data *Testing* (T)
- Pohon klasifikasi dibentuk dari L dengan penyeleksian menggunakan *test sample estimate*. Jalankan data T pada pohon klasifikasi tersebut dan akan menghasilkan tingkat kesalahan klasifikasi (*misclassification rate*) $e_s(L, T)$.
- Sampel Bootstrap L_B diambil dari L dan pohon klasifikasi dibangun dari L_B dengan *test sample estimate*. Ulangi langkah ini sebanyak M kali sehingga diperoleh pohon klasifikasi $\phi_1(x), \dots, \phi_M(x)$. Mengadopsi rekomendasi dari para peneliti terdahulu maka dalam penelitian ini replikasi sample bootstrap yang dibuat adalah 25, 50, 75, 100, 150 dan 200 kali.
- Jalankan data T pada pohon klasifikasi yang terbentuk dari masing-masing sampel bootstrap. Kemudian hitung bagging misclassification rate) $e_s(L_B, T)$.
- Pemilahan data secara acak diulang dengan iterasi 100 kali dan hitung \bar{e}_s dan \bar{e}_B yang merupakan rata-rata dari 100 iterasi.
- Membandingkan hasil klasifikasi CART dengan bagging CART.

3. Hasil dan Pembahasan

3.1 Klasifikasi CART

Hasil klasifikasi CART ketepatan waktu kelulusan mahasiswa STIKOM Bali tahun 2016 menghasilkan pohon klasifikasi diperlihatkan pada Gambar 1. Pohon klasifikasi dibentuk dari 7 variabel. Variabel yang digunakan sebagai pemilah pembentukan pohon klasifikasi CART dan paling menentukan ketepatan waktu lulusan secara berurutan adalah lama menyusun skripsi (LMS), Indeks Predikat Kumulatif (IPK) dan Program studi (PS).



Gambar 1. Pohon Klasifikasi

Berdasarkan Gambar 1, lulusan tepat waktu (TW) terbagi menjadi 4 simpul dengan karakteristik sebagai berikut:

- a. Lama menyusun skripsi kurang dari 3 bulan.
- b. Lama menyusun skripsi antara 3 sampai 7 bulan, dengan IPK kurang dari atau sama dengan 3,2 pada Program Studi (PS) Sistem Informasi.
- c. Lama menyusun skripsi kurang dari atau sama dengan 4 bulan dengan IPK lebih dari 3,2 pada Program Studi Sistem Informasi.
- d. Lama menyusun skripsi lebih dari 7 bulan, dengan IPK kurang atau sama dengan 3,5.

Sedangkan klasifikasi kelulusan dengan tidak tepat waktu terbagi menjadi 5 simpul dengan karakteristik sebagai berikut:

- a. Lama menyusun skripsi antara 3 sampai 7 bulan, dengan IPK kurang dari atau sama dengan 3,2 pada Program Studi (PS) Sistem Komputer.
- b. Lama menyusun skripsi kurang dari 4 bulan dengan IPK lebih dari 3,2 pada Program Studi (PS) Sistem Komputer.
- c. Lama menyusun skripsi lebih dari 4 bulan dengan IPK lebih dari 3,2
- d. Lama menyusun skripsi kurang atau sama dengan 10 bulan, dengan IPK lebih dari 3,5.
- e. Lama menyusun skripsi lebih dari 10 bulan dengan IPK lebih 3,5.

Berdasarkan pohon klasifikasi CART pada Gambar 1 dengan total jumlah data sampel sebanyak 890 mahasiswa, sebanyak 374 data sampel diprediksi tepat waktu sedangkan 31 lainnya diprediksi tidak tepat waktu atau disklasifikasikan. Sedangkan pada data sampel tidak tepat waktu, diprediksi 393 data sampel tepat diklasifikasikan tidak tepat waktu namun data sampel yang mengalami disklasifikasikan yaitu 92. Hasil klasifikasi yang dihasilkan diperlihatkan pada Tabel 2.

Tabel 2. Ketepatan Klasifikasi pada Pohon Optimal

Aktual	Prediksi		Total
	Tepat Waktu	Tidak Tepat Waktu	
Tepat Waktu	374	92	466
Tidak Tepat Waktu	31	393	424
Total	405	485	890

Hasil perhitungan akurasi klasifikasi dengan menggunakan $1 - \text{APER}$ pada pohon klasifikasi CART bernilai 0,8618 atau dengan kata lain, data sampel yang tepat diklasifikasikan secara keseluruhan sebanyak 86,18%. *Sensitivity* pohon klasifikasi CART bernilai 0,9369 yang artinya 93,69% data sampel ketepatan lulusan tepat diklasifikasikan tepat waktu. Adapun *specificity* bernilai 0,8026 yang artinya sebanyak 80,26% ketepatan lulusan tepat diklasifikasikan tidak tepat waktu. Hasil akurasi klasifikasi tersebut menunjukkan bahwa CART bagus dalam mengklasifikasikan ketepatan waktu lulusan mahasiswa, Akan tetapi akurasi klasifikasi tidak stabil pada data testing. Selanjutnya, dilakukan metode *bagging* digunakan untuk memperbaiki akurasi klasifikasi CART.

3.2 Klasifikasi dengan Bagging CART

Hasil klasifikasi Bagging CART pohon klasifikasi yang dilakukan pada waktu ketepatan lulusan mahasiswa STIKOM Bali menghasilkan 3 variabel yang digunakan sebagai pemilah pohon klasifikasi secara berurutan yaitu lama menyusun skripsi, IPK dan program studi. Analisis *bagging* CART menghasilkan sebanyak 407 data sampel ketepatan waktu lulusan diklasifikasikan tepat waktu sedangkan 62 lainnya tidak tepat waktu atau disklasifikasikan sebagai kelulusan tidak tepat waktu. Pada data sampel tidak tepat waktu, hanya 399 data sampel tepat diklasifikasikan tidak tepat waktu sedangkan 22 lainnya salah diklasifikasikan sebagai lulusan tepat waktu. Hasil klasifikasi *bagging* CART dapat dilihat pada Tabel 3:

Tabel 3. Klasifikasi Bagging CART

Aktual	Prediksi		Total
	Tepat waktu	Tidak tepat waktu	
Tepat waktu	407	62	469
Tidak tepat waktu	22	399	421
Total	429	461	890

Sama halnya dengan akurasi klasifikasi pada CART sebelumnya, acuan ukuran akurasi klasifikasi Bagging CART digunakan nilai $1 - \text{APER}$, *sensitivity* dan *specificity*. Hasil perhitungan menunjukkan bahwa $1 - \text{APER}$ klasifikasi Bagging CART bernilai 0,9056 atau dengan kata lain, data yang tepat diklasifikasikan secara keseluruhan sebanyak 90,56%. *Sensitivity* klasifikasi Bagging CART bernilai

0,9523 yang artinya 95,23% data sampel lulus tepat diklasifikasikan tepat waktu. Adapun *specificity* bernilai 0,8678 yang artinya 86,78% ketepatan lulus tepat diklasifikasi tidak tepat waktu. Hal tersebut menunjukkan bahwa Bagging CART lebih bagus dalam mengklasifikasikan waktu kelulusan.

4. Simpulan

Berdasarkan hasil dan pembahasan dapat diambil kesimpulan sebagai berikut:

- a. Penerapan pohon klasifikasi CART untuk klasifikasi ketepatan waktu lulus menunjukkan bahwa variabel yang digunakan sebagai pemilah pohon klasifikasi CART dan paling menentukan status waktu kelulusan adalah lama menyusun skripsi, IPK dan Program Studi. Berkaitan dengan akurasi klasifikasi, data sampel yang tepat diklasifikasikan secara keseluruhan sebesar 86,18%. Sedangkan akurasi prediksi pohon klasifikasi CART untuk kelulusan tidak tepat waktu sebesar 80,26%.
- b. Klasifikasi *bagging* CART menghasilkan akurasi klasifikasi sebesar 90,56% dengan akurasi prediksi pohon klasifikasi CART untuk kelulusan tidak tepat waktu sebesar 86,78%. Metode *bagging* menghasilkan akurasi klasifikasi yang lebih baik dibandingkan dengan pohon klasifikasi CART. Bagging CART mampu meningkatkan 1-APER pohon klasifikasi CART sebesar 4,38%.

Saran yang dapat diberikan dalam penelitian ini adalah memberikan perhatian khusus kepada mahasiswa dalam menyelesaikan/menyusun skripsi supaya dipersingkat, karena berdasarkan pohon klasifikasi CART dinyatakan bahwa lama menyusun skripsi menjadi salah satu alasan klasifikasi dalam status kelulusan mahasiswa tepat waktu atau tidak tepat waktu, hasil akurasi klasifikasi di atas perlu dibandingkan dengan metode *ensemble* lainnya

Daftar Pustaka

- [1] Komalasari, W.B. Metode Pohon Regresi untuk Eksplorasi Data dengan Peubah yang Banyak dan Kompleks. *Informatika Pertanian*. 2007; **Volume 16**. http://www.litbang.deptan.go.id/warta-ip/pdf-file/5.wieta_ipvol16-1-2007 [15 Februari 2008, Pukul 5. 38]
- [2] Roman Timofeev. Classification And Regression Trees (CART) Theory and Applications. Thesis. Berlin. Humboldt University. 2004.
- [3] Wezel, M.V. dan Potharst, R. Improved Customer Choice Predictions using Ensemble Methods. *European Journal of Operational Research*. 2007; **Vol. 181**, hal. 436-452.
- [4] Leo Breiman. 1994. Bagging Predictors. Technical Report No. 241. Department of Statistics, University of California, California.
- [5] Muhammad Jamal Muttaqin. Metode Ensemble pada CART Untuk Perbaikan Klasifikasi Kemiskinan di Kabupaten Jombang. Tesis. Surabaya. Institut Teknologi Sepuluh Nopember. 2013.
- [6] Aquarahma Margasari. Penerapan metode CART (Classification and Regression Trees) dan analisis regresi logistik Biner pada klasifikasi profil mahasiswa FMIPA Universitas Brawijaya. *Student Journal UB*. 2014; Vol 2 no 4: hal. 257 – 260.
- [7] Padmini, I.A.S., NLH Suciptawati dan Made Susilawati. Analisis Waktu Kelulusan Mahasiswa dengan Metode CHAID, (studi kasus: FMIPA Universitas Udayana). *E-jurnal Matematika*. 2012; Vol. 1, No.1: hal. 89-93.
- [8] I Ketut Putu Suniantara. Klasifikasi Ketidaktepatan Waktu Kelulusan Mahasiswa Stikom Bali Menggunakan Classification And Regression Trees (CART) Dan Regresi Logistik Biner. STIKOM Bali. Penelitian Internal Tahap II. 2016.