

ANALISIS KETEPATAN WAKTU LULUS MAHASISWA DENGAN MENGGUNAKAN BAGGING CART

Winalia Agwil^{1)*}, Herlin Fransiska²⁾, Nurul Hidayati³⁾
^{1,2,3}Jurusan Matematika, FMIPA, Universitas Bengkulu
Jl. W. R. Supratman, Bengkulu 38125

*winaliaagwil@unib.ac.id

Abstrak

Ketepatan waktu lulus mahasiswa menjadi salah satu indikator penilaian kelayakan program studi sebagai unit pelaksana pendidikan pada perguruan tinggi. Mengetahui faktor yang mempengaruhi waktu lulus mahasiswa akan membantu program studi dan dosen dalam mengambil keputusan untuk meningkatkan kuantitas mahasiswa lulus tepat waktu. Tujuan dari penelitian ini adalah untuk mendapatkan gambaran tentang karakteristik mahasiswa yang mempengaruhi ketepatan waktu lulus mahasiswa program studi S1 Matematika dengan menggunakan metode Ensemble Tree. Metode Ensemble tree yang digunakan adalah Bagging CART, dengan harapan dapat menghasilkan performa klasifikasi yang tinggi dan gambaran karakteristik mahasiswa yang baik. Data yang digunakan adalah data mahasiswa program studi S1 Matematika dari tahun 2010 sampai dengan 2019. Kebaikan klasifikasi dilihat dari nilai Accuracy, Sensitivity dan Specificity. Metode pohon klasifikasi tunggal (CART) memberikan nilai Accuracy sebesar 82.1% , sensitivity sebesar 68.2 % dan specificity sebesar 91.2 %. Sedangkan dengan metode Bagging CART diperoleh Accuracy sebesar 85.7% , sensitivity sebesar 77.3 % dan specificity sebesar 91.2 %. Berdasarkan perbandingan nilai akurasi yang diperoleh dapat disimpulkan bahwa menerapkan metode Bagging pada pohon klasifikasi tunggal CART dapat meningkatkan performa klasifikasi.

Kata Kunci: ketepatan waktu lulus, ensemble tree, bagging CART, performa klasifikasi.

PENDAHULUAN

Mahasiswa merupakan salah satu aspek penilaian keberhasilan maupun kelayakan suatu program studi dalam penyelenggaraan pendidikan tinggi. Indikator penilaian yang melibatkan mahasiswa diantaranya adalah perbandingan

jumlah pendaftar terhadap yang lulus seleksi, persentase mahasiswa yang melakukan pendaftaran ulang, rata-rata IPK mahasiswa, prestasi akademik maupun non-akademik yang diraih selama kuliah, dan juga persentase kelulusan tepat waktu (BAN-PT, 2019). Sehingga, Ketepatan

waktu lulus mahasiswa menjadi salah satu hal yang harus diperhatikan universitas maupun program studi sebagai unit pelaksana pendidikan tinggi.

Ketepatan waktu lulus mahasiswa memiliki kriteria yang berbeda untuk setiap program yang tersedia pada jenjang perguruan tinggi. Mahasiswa program D3 (Diploma) dikatakan lulus tepat waktu apabila dapat menyelesaikan studi kurang atau sama dengan tiga tahun dan dikategorikan lulus tidak tepat waktu apabila menyelesaikan studi lebih dari tiga tahun. Mahasiswa program S1 (Sarjana) dikatakan lulus tepat waktu apabila dapat menyelesaikan studi kurang atau sama dengan empat tahun dan dikategorikan lulus tidak tepat waktu apabila menyelesaikan studi lebih dari empat tahun. Demikian juga untuk mahasiswa program S2 (Magister) dikatakan lulus tepat waktu apabila dapat menyelesaikan studi kurang atau sama dua tahun jika lebih dari dua tahun dikategorikan tidak lulus tepat waktu dan program S3 (Doktor) jika dapat menyelesaikan studi kurang atau sama tiga tahun maka dapat dikategorikan lulus tepat waktu.

Banyak hal yang dapat mempengaruhi lama masa studi mahasiswa, baik faktor internal (minat, motivasi dan kemampuan mahasiswa) maupun faktor eksternal (lingkungan, daerah asal dan asal sekolah). Karakteristik mahasiswa yang berbeda menjadikan lama masa studi mahasiswa juga beragam. Hal ini menjadi menarik untuk dikaji sebagai bentuk upaya memperbaiki kualitas lulusan di masa yang akan datang dan juga dapat sebagai langkah antisipasi peningkatan jumlah kegagalan studi dalam hal ini, *drop out* (DO). Karakteristik mahasiswa yang memiliki kecenderungan lulus tepat waktu atau tidak dapat di analisis dengan menggunakan metode pohon klasifikasi. Beberapa metode

pohon klasifikasi yang sering digunakan antara lain adalah metode *Chi-squared Automatic Interaction Detection* (CHAID) *Classification and Regression Tree* (CART).

Metode CART telah banyak digunakan dalam analisis klasifikasi karena terbukti memberikan tingkat kesalahan klasifikasi yang kecil dan mudah dalam menginterpretasikan hasil analisis. Al-Amin dkk (2013) telah meneliti ketepatan kelulusan mahasiswa FEB dan Faperta dengan menggunakan metode CART, metode ini memberikan nilai akurasi 71% dengan karakteristik mahasiswa yang cenderung lulus tepat waktu dapat dilihat dari IPK dan jenis kelamin mahasiswa. Sebelumnya Pertiwi (2013) juga meneliti tentang analisis klasifikasi tepat lulus mahasiswa berdasarkan karakteristik mahasiswa dengan menggunakan metode CHAID, hasil penelitian menunjukkan bahwa akurasi 80%. Namun, hasil menunjukkan bahwa metode CHAID tidak baik ketika memprediksi kategori mahasiswa yang lulus tepat waktu yang mengindikasikan bahwa model klasifikasi yang terbentuk belum cukup baik. Banyak hal yang mempengaruhi ketidakstabilan pohon klasifikasi tunggal seperti CART dan CHAID, salah satunya jika terdapat ketidakseimbangan ukuran sampel antara kategori yang satu dan yang lainnya.

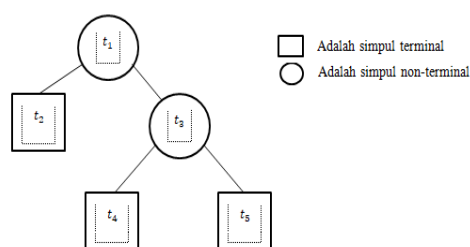
Untuk meningkatkan performa pohon klasifikasi tunggal dalam hal ini CART dapat dilakukan dengan menerapkan suatu metode *Ensemble*. Beberapa metode *Ensemble* yang dapat digunakan yakni *Bagging*, *Boosting* dan *Random Forest*. Menurut Breiman (1996) dari banyak percobaan yang dilakukan, salah satu metode *Ensemble* dalam hal ini metode *Bagging* mampu mengurangi tingkat kesalahan dalam pengklasifikasian. Selain itu, Hastie dkk (2008) berpendapat bahwa

metode Bagging ini dapat mengurangi kesalahan pengklasifikasian yang dihasilkan oleh model klasifikasi tunggal.

Setiap metode *Ensemble* memberikan kelebihan dan kekurangan tertentu, baik dalam hal kesederhanaan proses dan komputasi. Untuk tujuan meningkatkan performa klasifikasi model klasifikasi yang akan digunakan untuk menganalisis ketepatan waktu lulus mahasiswa dan mempertimbangkan kesederhanaan komputasi maka penulis tertarik untuk melakukan penggabungan metode Bagging dengan algoritma CART yang biasa disebut dengan Bagging CART.

CART merupakan salah satu metode pohon keputusan yang digunakan untuk variabel respon berupa numerik dan kategorik. Jika variabel respon berupa data kategorik dinamakan pohon klasifikasi sedangkan pada variabel respon numerik dinamakan pohon regresi. Pohon klasifikasi maupun pohon regresi terbentuk dari proses rekursif biner pada gugus data sehingga nilai variabel respon pada setiap gugus data hasil pemilahan akan lebih homogen (Breiman et al. 1984).

Ide dasar dari metode ini adalah menyeleksi variabel prediktor yang memiliki interaksi yang paling kuat dengan variabel responnya. Lewis (2020) menyebutkan metode ini sebagai klasifikasi *binary recursive partitioning* karena pada setiap simpul yang dihasilkan akan disekat atau dipecah menjadi dua simpul anak ydan dilakukan secara terus menerus sampai tidak dapat dipecah lagi. Gambar 1 merupakan visual metode CART dengan dua simpul non-terminal dan tiga simpul terminal.



Gambar 1. Struktur Pohon Algoritma CART

Andaikan S adalah gugus data yang akan digunakan untuk membangun pohon klasifikasi maupun pohon regresi maka langkah-langkah pada algoritma CART adalah sebagai berikut:

1. Identifikasi peubah bebas dan nilainya (atau levelnya jika peubah berupa data kategorik) yang dapat digunakan sebagai pemisah (splitting) data S menjadi dua atau lebih sub-set data.
2. Lakukan iterasi terhadap langkah 1 terhadap subset-subset yang ada sampai terpenuhi hal berikut:
 - a. Semua subset dengan nilai homogen
 - b. Tidak terdapat peubah bebas untuk dijadikan pemisah
 - c. Jumlah amatan didalam subset sudah terlalu sedikit untuk mendapatkan pemisahan yang memuaskan
3. Lakukan pemangkasan (*pruning*) jika pohon yang dihasilkan dinilai terlalu besar.

Tahapan awal seluruh gugus data berada pada simpul akar yang kemudian disekat menjadi dua anak gugus data, simpul kiri dan simpul kanan. Simpul yang memiliki nilai belum homogen atau bercampur disekat kembali sehingga dihasilkan simpul dengan nilai homogen, simpul akhir disebut juga simpul terminal, sedangkan simpul yang masih bias disekat disebut simpul non terminal.

Kriteria kebaikan sekat merupakan alat yang digunakan untuk melihat kebaikan sekat- s pada simpul t . jika sekat s pada simpul t menyekat data dengan proporsi p_R ke dalam simpul bagian kanan (t_R) dan proporsi p_L ke dalam simpul bagian kiri (t_L) maka kebaikan sekat didefinisikan sebagai berikut:

$$\Delta i(s, t) = i(t) - p_R i(t_R) - p_L i(t_L) \quad (1)$$

Dengan nilai keheterogenanan data pada setiap simpul diukur dengan formula berikut:

$$i(t) = -\sum_{j=1}^J p(j|t) \log_2 p(j|t) \quad (2)$$

$p(j|t)$ merupakan proporsi kategori j pada simpul t dan $p(k|t)$ merupakan proporsi kategori k pada simpul t . Nilai $i(t)$ berada pada selang nol sampai dengan satu, $i(t)$ bernilai maksimum ketika ketika kelas belum homogen dan sebaliknya akan bernilai minimum jika nilai kelas dalam simpul telah homogen.

Model klasifikasi tunggal (CART) dapat ditingkatkan akurasi klasifikasinya dengan menggunakan metode Ensembel salah satunya Bagging. Bagging merupakan singkatan dari *bootstrap aggregating*. *Bagging* adalah salah satu metode *ensemble* yang digunakan untuk meningkatkan performa klasifikasi (Breiman, 1996).

Jika dilihat dari namanya *bagging* terdiri dari dua tahapan yakni *bootstrap* yakni pengambilan contoh dari data yang tersedia (*resampling*) dan *aggregating* yaitu penggabungan banyak nilai dugaan menjadi satu nilai dugaan. Sehingga proses klasifikasi dengan *bagging tree* adalah sebagai berikut (Sartono dan Syafitri, 2010):

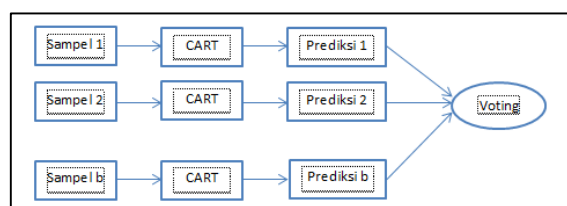
1. Tahapan *Bootstrap*

- a. Penarikan contoh acak sebanyak n contoh dengan pengembalian dari gugus data training.
- b. Susun data terbaik dengan menggunakan data tersebut.
- c. Lakukan pengulangan langkah a dan b sebanyak k kali sehingga diperoleh k buah pohon acak.

2. Tahapan *Aggregating*

Melakukan pendugaan gabungan berdasarkan k pohon yang telah terbentuk, pemilihan satu dugaan dapat menggunakan *majority vote*.

Selain *majority vote* atau pengambilan suara terbanyak, penggabungan dugaan dapat juga dilakukan dengan melakukan penjumlahan dugaan peluang masing-masing kelas, karena pada setiap dugaan pohon yang dihasilkan berupa peluang (Rokack, 2008). Secara umum proses Bagging dengan menggunakan algoritma CART sebagai model klasifikasinya dapat dilihat pada gambar 2.



Gambar 2. Ilustrasi Bagging CART

Performa klasifikasi metode CART dan *Bagging* CART dapat dihitung dari *Confusion Matrix*. *Confusion matrix* merupakan tabulasi dari ketepatan klasifikasi pada data prediksi dan data aktual. *True Positive* adalah jumlah amatan yang tepat klasifikasi dari kelas positif, *True Negative* adalah jumlah amatan yang tepat klasifikasi dari kelas negatif, *False Negative* adalah jumlah amatan yang salah klasifikasi dari kelas positif sedangkan *False Positive*

adalah jumlah amatan yang salah klasifikasi dari kelas negatif.

Tabel 1. *Confusion Matrix*

Prediksi	Aktual	
	Positif	Negatif
Positif	<i>True Positive</i> (TP)	<i>False Positive</i> (FP)
Negatif	<i>False Negative</i> (FN)	<i>True Negative</i> (TN)

Berdasarkan nilai pada *Confusion Matrix* dapat dihitung nilai *Accuracy*, *Sensitivity* dan *Specificity*. *Accuracy* menyatakan tingkat ketepatan *classifier* dalam mengklasifikasi amatan. Berikut adalah formula yang digunakan dalam melihat performa klasifikasi :

$$Accuracy = \frac{TP+TN}{(TP+TN+FN+FP)} \quad (3)$$

$$Sensitivity = \frac{TP}{TP+FN} \quad (4)$$

$$Specificity = \frac{TN}{TN+FP} \quad (5)$$

Tingkat ketepatan klasifikasi juga dapat dilihat nilai AUC, yaitu luas di bawah kurva ROC yang berkisar antara 0 hingga 1. Kurva ROC adalah kurva yang menggambarkan performa klasifikasi secara dua dimensi (Fawcett, 2006). Kurva tersebut adalah plot persentase *False Positive* (1-*Specificity*) dengan persentase *True Positive*.

METODE PENELITIAN

Penelitian ini memanfaatkan rekam data akademik dan non akademik mahasiswa program studi S1 Matematika FMIPA Universitas Bengkulu. Data mahasiswa yang digunakan adalah mahasiswa dengan periode kelulusan tahun 2015 sampai dengan 2019. Variabel yang digunakan dalam penelitian antara lain, Ketepatan waktu lulus mahasiswa yang dikategorikan menjadi dua yakni “Tepat” dan “Tidak Tepat”, IPK, Jenis Kelamin, Daerah Asal, Asal Sekolah, Jalur

Masuk PT, Status Pembiayaan Kuliah dan Status Keluarga.

Sebagaimana yang telah dijelaskan sebelumnya bahwa akan dilakukan pemodelan klasifikasi dengan menggunakan dua metode yakni CART dan Bagging CART. Langkah awal yang dilakukan adalah mempartisi atau membagi dataset menjadi dua bagian yaitu data *training* dan data *testing* dengan rasio 80:20. Data *training* digunakan untuk pemodelan sedangkan data *testing* digunakan untuk mengevaluasi performa klasifikasi atau model yang telah terbentuk.

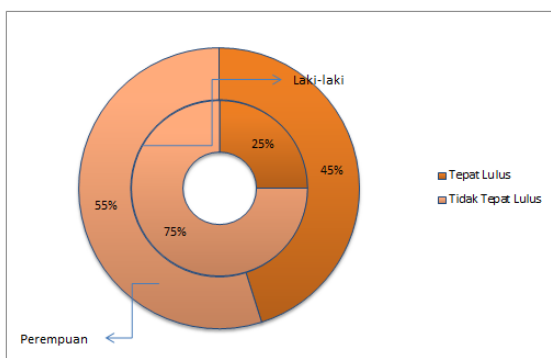
Selanjutnya melakukan pemodelan klasifikasi dengan menggunakan metode CART memanfaatkan data *training*. Proses pembentukan metode klasifikasi akan berhenti apabila banyak objek pengamatan dalam simpul terminal sebanyak 11 pengamatan. Setelah terbentuk model klasifikasi maka langkah selanjutnya adalah melakukan prediksi dengan memanfaatkan data *testing* dan membandingkan hasil prediksi dengan data aktual. Kebaikan model klasifikasi yang terbentuk dapat dilihat dari nilai *Accuracy*, *Sensitivity* dan *Specificity*.

Membentuk model klasifikasi metode Bagging CART dapat dilakukan dengan melakukan resampling sebanyak *b* kali. Kemudian dari setiap sampel yang terbentuk dibangun model klasifikasi CART. Dengan memanfaatkan prediksi yang dihasilkan setiap model CART yang terbentuk dari *b* sampel, hasil prediksi yang dipilih merupakan kategori terbanyak. Selanjutnya sama halnya dengan metode CART yakni membandingkan hasil prediksi dan data aktual kemudian kebaikan model Bagging CART dapat dilihat dari nilai *Accuracy*, *Sensitivity* dan *Specificity*. serta hal-hal lain yang berkaitan dengan cara penelitian.

Bagian ini dapat dibagi menjadi beberapa sub bab, tetapi tidak perlu mencantumkan penomorannya.

HASIL DAN PEMBAHASAN

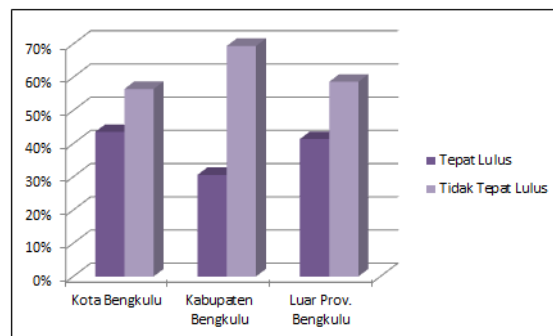
Data yang digunakan dalam penelitian ini adalah mahasiswa dengan periode kelulusan tahun 2015-2019 yakni sebanyak 284 mahasiswa. Proporsi mahasiswa yang lulus tepat waktu pada periode tersebut adalah sebanyak 39.4%. Jika ditinjau dari jenis kelamin mahasiswa, mahasiswa perempuan cenderung lebih banyak lulus tepat waktu yakni sebesar 45% dari total jumlah mahasiswa perempuan. Sedangkan mahasiswa laki-laki hanya 25% dari total mahasiswa laki-laki yang lulus tepat waktu, hal ini dapat dilihat dari Gambar 3.



Gambar 3. Persentase Ketepatan Waktu Lulus Mahasiswa Menurut Jenis Kelamin

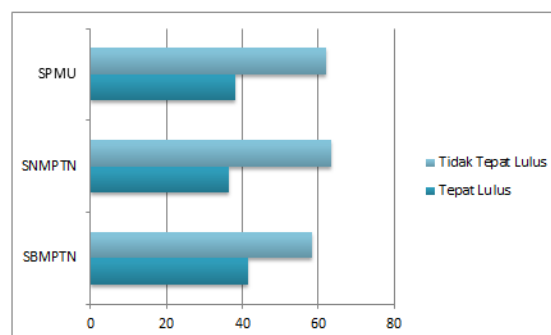
Dilihat dari daerah asal mahasiswa, mahasiswa yang berasal dari kota Bengkulu cenderung lebih tepat lulus dibandingkan mahasiswa yang berasal dari kabupaten dalam Provinsi Bengkulu dan luar Provinsi Bengkulu. Hal ini dapat dilihat dari Gambar 3 yang menampilkan proporsi ketepatan lulus mahasiswa yang berasal dari Kota Bengkulu sebesar 44% dari total 170 mahasiswa yang berasal dari Kota Bengkulu. Sedangkan, untuk mahasiswa yang berasal dari kabupaten yang berada di

Provinsi Bengkulu berjumlah 85 orang dengan proporsi yang lulus tepat waktu sebesar 30.6% dan mahasiswa yang berasal dari luar Provinsi Bengkulu sebanyak 29 mahasiswa dengan proporsi ketepatan lulus sebesar 41.4%.



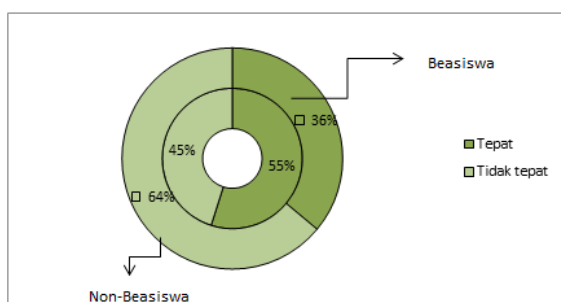
Gambar 4. Persentase Ketepatan Waktu Lulus Mahasiswa Menurut Daerah Asal

Terdapat tiga jalur masuk mahasiswa Universitas Bengkulu yakni, SNMPTN, SBMPTN dan SPMU. Jumlah mahasiswa dengan jalur masuk SNMPTN sebanyak 159, SBMPTN 104 dan SPMU sebanyak 21 mahasiswa. Jika dilihat dari Gambar 5, mahasiswa dengan jalur masuk melalui SBMPTN memiliki proporsi lulus tepat waktu lebih tinggi dibandingkan jalur lain yakni sebanyak 41.5% yang disusul oleh mahasiswa dengan jalur masuk universitas melalui SBMPTN sebanyak 36.5% dan mahasiswa dengan jalur masuk universitas melalui SPMU sebanyak 38.1%.



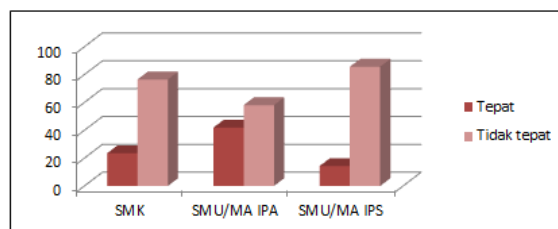
Gambar 5. Persentase Ketepatan waktu Lulus Mahasiswa Menurut Jalur Masuk Universitas Bengkulu

Gambar 6 meyajikan persentase ketepatan lulus mahasiswa menurut pembiayaan perkuliahan yang terbagi atas biaya kuliah dengan beasiswa dan non-beasiswa. Dari gambar terlihat bahwa mahasiswa yang dibiayai melalui beasiswa lebih cenderung memiliki ketepatan lulus yang lebih tinggi dibandingkan mahasiswa non-beasiswa, hal ini terlihat dari proporsi tepat lulus mahasiswa beasiswa sebesar 55% dari total mahasiswa dengan pembiayaan beasiswa. Sedangkan mahasiswa non-beasiswa memiliki proporsi tepat lulus sebesar 36 % dari total mahasiswa dengan pembiayaan non-beasiswa.



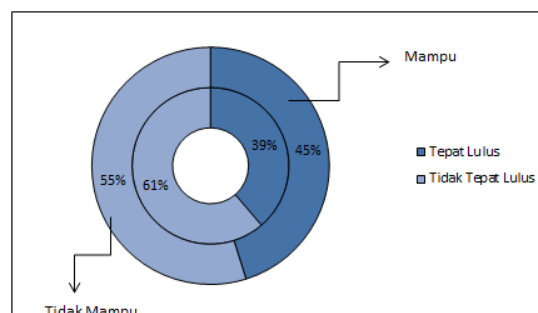
Gambar 6. Persentase Ketepatan Waktu Lulus Mahasiswa Menurut Jenis Pembiayaan Kuliah

Ditinjau dari asal sekolah mahasiswa yang dalam hal ini dibagi menjadi, SMU/MA IPA, SMU/MA IPS dan SMK. Mahasiswa yang berasal dari SMU/MA IPA memiliki kecenderungan untuk lulus tepat waktu dibandingkan mahasiswa yang berasal dari SMU/MA IPS dan SMK. Hal ini dapat dilihat dari Gambar 7, proporsi ketepatan lulus mahasiswa yang berasal dari SMU/MA IPA sebesar 41.9% dari total 253 mahasiswa sedangkan mahasiswa yang berasal dari SMU/MA IP dan SMK berturut-turut sebesar 14.3% dan 23.5%.



Gambar 7. Persentase Ketepatan Waktu Lulus Mahasiswa Menurut Asal Sekolah

Jika dilihat dari status keuangan keluarga, mahasiswa yang tergolong mampu lebih cenderung menyelesaikan studi tepat waktu dibandingkan dengan mahasiswa yang tergolong tidak mampu, dengan proporsi sebesar 45% dari total mahasiswa tergolong mampu. Sedangkan proporsi ketepatan waktu lulus mahasiswa tidak mampu sebesar 39% dari total mahasiswa tidak mampu.



Gambar 8. Persentase Kelulusan Tepat Waktu Menurut Status Keuangan Keluarga

Berdasarkan ringkasan data IPK yang terlampir pada Tabel 2, dapat disimpulkan bahwa mahasiswa yang menyelesaikan studi tepat waktu memiliki rata-rata IPK sebesar 3.25 dengan simpangan baku sebesar 0.23. sedangkan mahasiswa yang tergolong tidak tepat waktu dalam menyelesaikan studi memiliki rata-rata IPK sebesar 2.77 dengan simpangan baku sebesar 0.32.

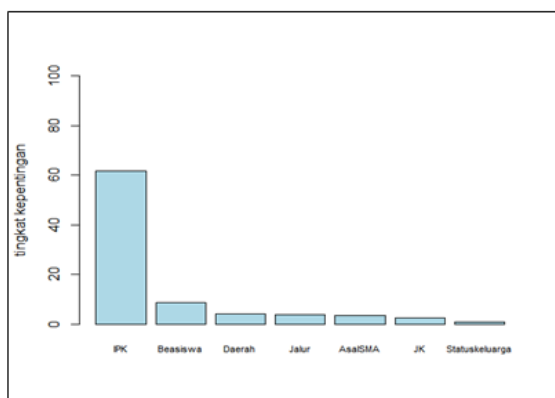
Tabel 2. Ringkasan IPK Mahasiswa

	Rata-Rata IPK	SD	Total
Tepat	3.25	0.23	112
Tidak Tepat	2.77	0.32	172

Model Klasifikasi Menggunakan CART

Sebagaimana yang telah dibahas pada metode penelitian, bahwa dalam pembentukan pohon klasifikasi CART terlebih dahulu dilakukan pembagian data menjadi 2 bagian yakni data training dan data testing dengan proporsi 80: 20. Proporsi pembagian data training dan data testing dapat dilakukan dengan beberapa opsi, namun harus diperhatikan bahwa data training harus memiliki proporsi yang lebih besar dibandingkan data testing (Breiman 1984 dalam Duda et.al 2000).

Sesuai dengan tujuan penelitian yakni mengetahui karakteristik mahasiswa yang mempengaruhi ketepatan waktu lulus mahasiswa program studi S1 Matematika FMIPA Universitas Bengkulu. Metode yang digunakan adalah pohon klasifikasi CART, yang mana model dibangun menggunakan data training. Urutan tingkat kepentingan variabel dapat dilihat dari Gambar 9.



Gambar 9. Tingkat Kepentingan Variabel dengan CART

Terlihat bahwa IPK memiliki peran paling besar dalam mengklasifikasikan ketepatan waktu lulus mahasiswa dengan nilai kepentingan 61.7. Disusul variabel

Beasiswa dengan tingkat kepentingan diurutan kedua dengan nilai 8.6. Diurutan ketiga, keempat dan kelima adalah variabel Daerah Asal, Jalur Masuk Universitas dan Asal SMA. Urutan paling bawah adalah Status Keuangan Keluarga. Dapat disimpulkan bahwa yang paling memiliki peran penting dalam penentuan mahasiswa siswa dapat lulus tepat waktu atau tidak adalah IPK.

Selain melihat karakteristik mahasiswa yang mempengaruhi ketepatan waktu lulus, penelitian ini juga bertujuan untuk membangun model pohon klasifikasi CART dalam mengklasifikasikan ketepatan waktu lulus mahasiswa. Kebaikan suatu model klasifikasi dapat dilihat dari nilai accuracy, sensitivity dan specificity. Pengujian model pohon klasifikasi yang terbentuk dilakukan dengan cara mengaplikasikan data testing pada model pohon klasifikasi yang telah dibangun sebelumnya dengan menggunakan data training. Hasil klasifikasi metode CART disajikan pada Tabel 3.

Tabel 3. Hasil Klasifikasi CART

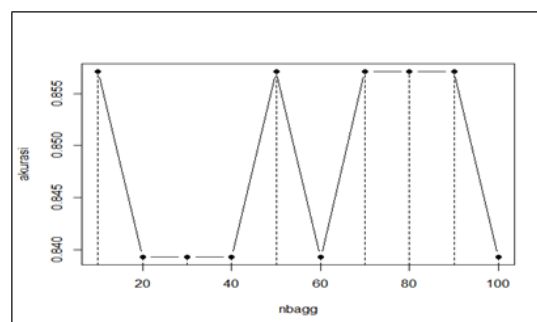
Prediksi	Aktual	
	Tepat	Tidak Tepat
Tepat	15	3
Tidak Tepat	7	31

Berdasarkan tabel diatas dapat dilihat bahwa terdapat 7 observasi yang diklasifikasikan menjadi mahasiswa yang lulus tidak tepat waktu seharusnya merupakan mahasiswa yang lulus tepat waktu. Selain itu, terdapat 3 observasi yang diklasifikasikan menjadi mahasiswa yang lulus tepat waktu seharusnya merupakan mahasiswa yang lulus tidak tepat waktu. Sehingga nilai Accuracy yang diperoleh adalah sebesar 82.1%, nilai Accuracy menggambarkan bahwa secara keseluruhan ketepatan klasifikasi yang dihasilkan model

CART adalah 82.1%. Nilai *sensitivity* sebesar 68.2 %, menggambarkan bahwa tingkat keberhasilan klasifikasi mahasiswa yang lulus tepat waktu adalah 68.2% dan nilai *specificity* sebesar 91.2 % menggambarkan bahwa tingkat keberhasilan klasifikasi mahasiswa yang lulus tidak tepat waktu sebesar 91.2%.

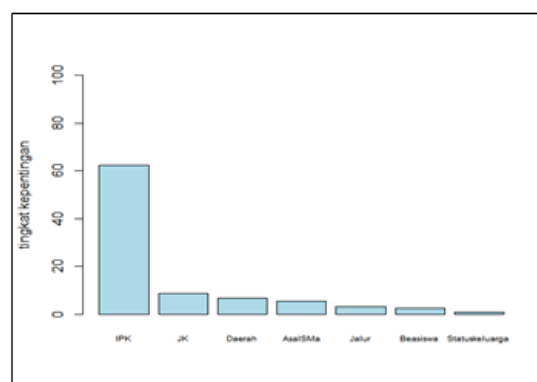
Model Klasifikasi Menggunakan Bagging CART

Pada bagian sebelumnya, telah dijelaskan bahwa untuk meningkatkan performa klasifikasi pada model pohon klasifikasi tunggal (CART) dapat diterapkan metode Ensemble salah satunya Bagging. Algoritma Bagging CART memiliki ide dasar yakni membentuk beberapa pohon klasifikasi baru dengan memanfaatkan data sampel baru yang diperoleh dari proses resampling. Gambar 10 menunjukkan beberapa resampling yang dilakukan dan nilai akurasi model yang dihasilkan. Dari Gambar 10 dapat dilihat bahwa resampling yang memiliki nilai akurasi sama yakni ketika diulang sebanyak 20, 50, 70, 80 dan 90 kali yakni dengan tingkat akurasi lebih dari 85.5%. Akurasi atau ketepatan klasifikasi yang digambarkan pada gambar 9 diperoleh dari proses majority vote yakni mengambil suara terbanyak dari nilai prediksi yang dihasilkan dari setiap pohon klasifikasi tunggal yang terbentuk. karena beberapa kali perulangan pengambilan sampel menghasilkan akurasi yang sama maka penulis memutuskan untuk menggunakan melakukan resampling sebanyak 50 kali.



Gambar 10. Nilai Akurasi Berdasarkan Banyak Resampling

Resampling yang dilakukan sebanyak 50 kali tentunya menghasilkan beragam pohon klasifikasi tunggal dengan tingkat kepentingan variabel independen yang berbeda pada setiap pohon klasifikasinya. Namun jika secara rata-rata berikut adalah variabel independen yang memiliki tingkat kepentingan dalam pembentukan klasifikasi ketepatan waktu lulus mahasiswa (Gambar 11).



Gambar 11. Tingkat Kepentingan Variabel Independen Bagging CART

Berdasarkan gambar 11 dapat dilihat bahwa Indeks Prestasi Kumulatif yang sering disingkat dengan IPK merupakan variabel yang memiliki peran penting dalam membangun pohon klasifikasi. Hal ini menunjukkan bahwa IPK merupakan faktor dominan dalam mencirikan ketepatan waktu lulus mahasiswa. Mahasiswa dengan IPK yang lebih tinggi memiliki peluang yang besar untuk menyelesaikan studi tepat

waktu. Sedangkan untuk variabel independen yang lain memiliki tingkat kepentingan yang hampir sejajar atau sama.

Sama halnya dengan metode CART yang telah dijelaskan pada pembahasan sebelumnya, selain melihat karakteristik mahasiswa yang mempengaruhi ketepatan waktu lulus, penelitian ini juga bertujuan untuk membangun model pohon klasifikasi Bagging CART dalam mengklasifikasikan ketepatan waktu lulus mahasiswa. Kebaikan suatu model klasifikasi dapat dilihat dari nilai *accuracy*, *sensitivity* dan *specificity*. Pengujian model pohon klasifikasi yang terbentuk dilakukan dengan cara mengaplikasikan data testing pada model pohon klasifikasi yang telah dibangun sebelumnya dengan menggunakan data training. Hasil klasifikasi metode Bagging CART disajikan pada Tabel 4.

Tabel 4. Hasil Klasifikasi Bagging CART

Prediksi	Aktual	
	Tepat	Tidak Tepat
Tepat	17	3
Tidak Tepat	5	31

Berdasarkan tabel diatas dapat dilihat bahwa terdapat 5 observasi yang diklasifikasikan menjadi mahasiswa yang lulus tidak tepat waktu seharusnya merupakan mahasiswa yang lulus tepat waktu. Dari jumlah kesalahan klasifikasi pada kategori “Tepat” dapat dilihat bahwa terjadi peningkatan akurasi klasifikasi yang sebelumnya terdapat 7 observasi yang salah klasifikasi pada kategori “Tepat”. Sedangkan pada kategori “Tidak Tepat” kesalahan klasifikasinya sama dengan metode CART yakni 3 observasi. Sehingga nilai *Accuracy* yang diperoleh adalah sebesar 85.71%, nilai *Accuracy* menggambarkan bahwa secara keseluruhan ketepatan klasifikasi yang dihasilkan model Bagging CART adalah 85.71%. Nilai

sensitivity sebesar 77.3 %, menggambarkan bahwa tingkat keberhasilan klasifikasi mahasiswa yang lulus tepat waktu adalah 71.3% dan nilai *specificity* sebesar 91.2 % menggambarkan bahwa tingkat keberhasilan klasifikasi mahasiswa yang lulus tidak tepat waktu sebesar 91.2%.

Perbandingan Klasifikasi Metode CART dan Bagging CART

Ketepatan klasifikasi mahasiswa yang termasuk kategori lulus tepat waktu atau tidak tepat waktu merupakan ukuran kebaikan pohon klasifikasi yang terbentuk. Analisis klasifikasi ketepatan waktu lulus mahasiswa dengan menggunakan metode CART dan *Bagging* CART memberikan hasil akurasi yang berbeda. Dibanding metode pohon klasifikasi tunggal (CART), penerapan metode *Bagging* pada pohon klasifikasi CART mampu meningkatkan ketepatan klasifikasi total (akurasi) sebesar 3.57% dan mampu meningkatkan ketepatan klasifikasi pada kategori lulus tepat waktu yang merupakan data minoritas yakni sebesar 9.1%. Sedangkan ketepatan klasifikasi kategori lulus tidak tepat waktu dapat dilihat dari nilai *specificity*-nya yakni 91.2% yang sama pada kedua metode. Dengan demikian dapat disimpulkan bahwa metode Bagging CART lebih baik dari metode CART tunggal dalam mengklasifikasikan ketepatan waktu lulus mahasiswa S1 Matematika FMIPA Universitas Bengkulu.

Nilai akurasi dengan menggunakan metode *Bagging* CART lebih baik daripada metode klasifikasi tunggal CART karena pada metode Bagging CART dilakukan resampling sebanyak 50 kali dan dilakukan pemodelan kemudian dilakukan prediksi dari masing-masing sampel yang terbentuk. Pemilihan prediksi yang mewakili adalah dengan menggunakan *majority vote* atau

memilih suara terbanyak. Dengan adanya pengulangan pengambilan sampel sebanyak 50 kali ini mengakibatkan adanya kekonvergenan hasil prediksi.

Tabel 5. Perbandingan Ketepatan Klasifikasi CART dan Bagging CART

Metode	Sensitivity	Specificity	Akurasi
CART	68.2%	91.2%	82.14%
Bagging CART	77.3 %	91.2%	85.71%

SIMPULAN

Berdasarkan hasil yang telah dibahas sebelumnya dapat disimpulkan bahwa dengan menggunakan metode CART, ketepatan waktu lulus mahasiswa sangat dipengaruhi oleh IPK, selain itu juga dipengaruhi oleh jenis pembiayaan kuliah dan daerah asal. Jika menggunakan metode Bagging CART, ketepatan waktu lulus mahasiswa dipengaruhi oleh IPK dan jenis kelamin mahasiswa.

Selain melihat karakteristik mahasiswa yang mempengaruhi ketepatan waktu lulus, dari penelitian ini juga dapat dilihat bahwa metode Bagging CART dapat meningkatkan performa klasifikasi. Hal ini dapat dilihat dari peningkatan nilai akurasi total maupun nilai *sensitivity*-nya. Perbaikan nilai akurasi dikarenakan adanya tahapan resampling dan kemudian prediksi diperoleh melalui proses *majority vote* yang memberikan kekonvergenan hasil prediksi.

UCAPAN TERIMA KASIH

Terima kasih penulis sampaikan kepada Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Bengkulu yang telah membiayai penelitian ini.

DAFTAR PUSTAKA

- Al-Amin, F.N, Indahwati, dan Anggraini, Y. 2013. *Analisis Ketepatan Waktu Lulus Berdasarkan Karakteristik Mahasiswa Di FEM dan Faperta menggunakan metode CART*. Xplore Vol 2(1). IPB.
- Badan Akreditasi Nasional Perguruan Tinggi. 2019. *Naskah Akademik IAPT 3.0*. BAN-PT: Jakarta.
- Breiman L, Friedman J. H, Olshen R. A, Stone CJ. 1984. *Classification and Regression Trees*. New York: Chapman & Hall/CRC.
- Breiman L. 1996. *Bagging Predictors*. *Machine Learning* 24: 123–140.
- Duda, R. O., Hart, P. E., Stork, D. G. 2000. *Pattern Classification*. California: Department at San Jose State University.
- Fawcett T. 2006. *An introduction to ROC analysis*. *Pattern Recognition Letters*. 27: 861–874.
- Lewis, R. J. 2020. *An Introduction to Classification and Regression Trees (CART) Analysis*. Presented at 2000 Annual Meeting of Society for Academic Emergency Medicine of San Francisco. California
- Pertiwi, R.A, Indahwati, dan Afendi, F.M. 2013. *Analisis CHAID untuk Identifikasi Ketepatan Waktu Lulus Berdasarkan Karakteristik Mahasiswa*. Xplore Vol 2(1). IPB.
- Rokach, L. 2008. *Ensemble Methods For Classifiers dalam Data Mining and Knowledge Discovery Handbook (editor Maimon O. and Rokach L.)*. New York: Springer Science + Business Media.

Sartono, B. dan Syafitri., T. U. 2010. *Metode Pohon Gabungan: Solusi Pilihan Untuk Mengatasi Kelemahan Pohon*

Regresi Dan Klasifikasi Tunggal. Forum Statistika dan Komputasi. P1-7.