

Klasifikasi Sekolah dalam Penerimaan Mahasiswa Baru Vokasi IPB Jalur USMI Menggunakan Metode CART

Erlinda Widya Widjanarko¹, Utami Dyah Syafitri^{1±}, Aam Alamudi¹

¹Department of Statistics, IPB University, Indonesia

‡corresponding author: utamids@apps.ipb.ac.id

Abstract

The selection of new student admissions for the IPB vocational school consists of several routes, one of which is the USMI route. In order to improve its performance, it is necessary to evaluate the USMI new student admission system. Previously, research with the same objective had been carried out using the clustering method. The study resulted in three clusters in which schools were differentiated based on commitment and quality. This study aims to create a classification model of the clusters obtained using the CART method. Classification and Regression Tree (CART) is a nonparametric classification technique that produces a single decision tree. The CART method can involve mixed-type data. The classification model generated from the 2019 data yields an accuracy of 98.52%. However, the results of 2019 model evaluation with the 2020 data are still not good enough to predict with an accuracy of 57.22%, so the 2020 data is re-clustered and produces three clusters. Furthermore, the classification model was remade with 2020 data, resulting in an accuracy of 97.47%. However, the results of the 2020 model evaluation with the 2021 data are still not good enough to predict with an accuracy of 44.34%, so the clustering in the previous year cannot be used for predictions of the following year's data. The grouping of schools for USMI applicants needs to be done by grouping schools every year.

Keywords : CART, classification, new student admissions, USMI, vocational

1. Pendahuluan

Sekolah vokasi IPB merupakan program diploma yang ditetapkan dan dibuka pada tahun 2017 melalui Surat Keputusan Rektor. Dalam penerimaan mahasiswa baru, Sekolah Vokasi IPB menyediakan tiga jalur masuk yang berbeda, yaitu Jalur Beasiswa Utusan Daerah (BUD), Reguler (ujian tulis), dan Undangan (USMI). Jalur USMI merupakan jalur undangan, di mana Sekolah Vokasi IPB akan mengirimkan surat kepada SMA/ sederajat sebagai undangan untuk mendaftarkan siswa nya di Sekolah Vokasi IPB, kemudian setiap sekolah akan mengirimkan nilai rapor dan beberapa berkas lain mengenai siswa yang didaftarkan kepada pihak IPB.

* Received: Aug 2022; Reviewed: Sep 2022; Published: Sep 2022

Dalam proses seleksinya, pihak IPB tidak hanya melihat dari nilai rapor saja, hal ini dikarenakan kriteria nilai rapor di setiap sekolah memiliki perbedaan. Oleh karena itu, pihak IPB juga memperhatikan beberapa aspek lainnya.

Dalam menerima mahasiswa baru, IPB university perlu melakukan evaluasi terhadap sistem yang ada. Hal ini bertujuan untuk meningkatkan kualitas dan efektivitas dalam penerimaan mahasiswa baru. Penelitian sebelumnya telah dilakukan oleh Sahili (2021). Dalam penelitian tersebut, sekolah-sekolah SMA/Sederajat dikelompokkan menjadi tiga gerombol berdasarkan komitmen dan kualitas sekolah-sekolah tersebut. Gerombol pertama berisi sekolah-sekolah yang memiliki komitmen dan kualitas yang tinggi, gerombol kedua berisi sekolah-sekolah yang memiliki komitmen dan kualitas tidak lebih tinggi dari gerombol pertama namun masih lebih tinggi dari gerombol ketiga, dan gerombol ketiga berisi sekolah-sekolah yang memiliki komitmen dan kualitas yang rendah dalam penerimaan mahasiswa baru. Penelitian tersebut mengevaluasi sistem penerimaan mahasiswa melalui analisis gerombol dengan algoritma *K-prototypes*.

Menurut Huang Z (1998), *K-prototypes* merupakan pengembangan dari algoritma *K-means* dan *K-modes*. Algoritma *K-means* hanya dapat digunakan pada peubah numerik yang kontinu saja, sehingga Huang (1998) memperkenalkan dua algoritma baru. Algoritma pertama merupakan algoritma *K-modes* yang dapat menangani data kategorik saja dan algoritma kedua adalah algoritma *K-Prototypes* yang dapat menangani data kategorik dan numerik. Algoritma *K-Prototypes* merupakan salah satu metode penggerombolan yang dapat menangani data bertipe campuran, seperti halnya dengan algoritma *two step clustering*. Namun, *K-prototypes* memiliki keunggulan yaitu algoritma yang tidak terlalu kompleks dibandingkan dengan algoritma *two step clustering* (Sulastri 2021).

Penelitian ini bertujuan untuk melakukan evaluasi terhadap sistem penerimaan mahasiswa baru jalur USMI dengan metode lebih lanjut, yaitu mengklasifikasikan sekolah-sekolah yang ada ke dalam tiga kelompok yang sudah ditemukan oleh peneliti sebelumnya. Menurut Han *et al.* (2012), Analisis klasifikasi merupakan analisis data yang menghasilkan suatu model atau pengklasifikasi untuk memprediksi label kelas pada setiap amatan. Analisis klasifikasi termasuk ke dalam *supervised learning* karena setiap amatan yang digunakan pada proses pelatihan sudah diketahui label kelas nya. Terdapat beberapa metode yang dapat digunakan dalam analisis klasifikasi seperti SVM, random forest, dan naïve bayes. Namun, model yang dihasilkan dari metode tersebut kurang bisa diinterpretasikan dengan baik. Salah satu bentuk keputusan dari analisis klasifikasi yang dapat diinterpretasikan dengan baik adalah pohon keputusan. Terdapat beberapa algoritma yang dipakai untuk menghasilkan pohon keputusan tersebut, beberapa di antaranya adalah CHAID, C5.0, dan CART. Menurut penelitian yang dilakukan oleh Pratiwi (2020), yang membandingkan antara algoritma C5.0 dan CART, didapatkan hasil bahwa akurasi algoritma CART lebih tinggi daripada akurasi

dari algoritma C5.0. Penelitian sebelumnya juga dilakukan oleh Aritonang *et al.* (2016) yang membandingkan algoritma CART dan CHAID, didapatkan hasil bahwa algoritma CART memiliki akurasi yang lebih tinggi dibandingkan dengan algoritma CHAID.

CART adalah metodologi statistik nonparametrik yang dikembangkan untuk topik analisis klasifikasi, baik untuk peubah yang bersifat kategorik maupun numerik (Breiman *et al.* 1993). Hasil akhir dari metode ini merupakan pohon keputusan. Penelitian menggunakan algoritma CART pernah dilakukan oleh Komalasari (2007). Menurutnya, metode ini dapat melibatkan data yang kompleks, dalam artian berdimensi besar ataupun bertipe data campuran. Sinambela (2008) yang dalam penelitiannya menggunakan metode CART juga berkesimpulan bahwa metode ini dapat melakukan eksplorasi data yang melibatkan peubah yang besar dan kompleks. Oleh karena itu, penelitian ini akan menghasilkan model untuk mengklasifikasikan apakah sebuah sekolah termasuk ke dalam kelompok tertentu menggunakan metode CART, yang selanjutnya akan menjadi pertimbangan pihak panitia penerimaan mahasiswa baru.

2. Metodologi

2.1 Data

Data yang digunakan untuk penelitian ini adalah data sekunder yang diperoleh dari Direktorat Administrasi Pendidikan dan Penerimaan Mahasiswa Baru (DAPPMB) IPB. Data sekolah tahun 2019 akan dijadikan sebagai data untuk membangun model awal. Sedangkan data tahun 2020 dan 2021 akan digunakan sebagai data untuk melakukan evaluasi model. Peubah yang digunakan pada penelitian ini merujuk pada penelitian yang dilakukan oleh Sahili (2021) dengan rincian seperti pada Tabel 1.

Tabel 1 Peubah yang digunakan pada penelitian

Notasi	Nama Peubah	Keterangan
Y	Komitmen dan kualitas sekolah	Terbagi menjadi tiga gerombol
X1	Asal provinsi	Terbagi menjadi 33 provinsi
X2	Jenis sekolah	1 = SMA, 2 = MA, 3 = SMK
X3	Akreditasi sekolah	1 = Akreditasi A, 2 = Akreditasi B 3 = Akreditasi C
X4	Status sekolah (negeri atau swasta)	1 = Negeri, 2 = Swasta
X5	Kategori prestasi akademik tiap sekolah	1 = Kategori A, 2 = Kategori B, 3 = Kategori C, 4 = Kategori D
X6	Rata-rata nilai rapor IPA tiap sekolah	Numerik
X7	Rata-rata nilai rapor IPS tiap sekolah	Numerik
X8	Persentase siswa lolos USMI IPB tiap sekolah	Numerik
X9	Persentase siswa daftar ulang USMI IPB tiap sekolah	Numerik
X10	Persentase mahasiswa dengan IPK \geq 3,00 tiap sekolah	Numerik

2.2 Metode Penelitian

Pada proses analisis data, terdapat dua alat atau perangkat lunak yang digunakan. Perangkat lunak yang pertama adalah *Rstudio* versi 4.0.2 dengan package yang digunakan adalah *rpart*, *caret*, dan *clustMixType*. Perangkat lunak yang kedua adalah

Microsoft excel 2013. Adapun tahapan analisis data pada penelitian ini adalah sebagai berikut.

1. Melakukan praproses data yang terdiri dari tahapan berikut.
 - a) Memisahkan data sekolah yang mendaftar USMI berdasarkan tahun akademik yaitu tahun 2019 sebagai data training, sedangkan data 2020 dan 2021 sebagai data testing.
 - b) Melakukan kategorisasi jenis sekolah yaitu SMA, MA, dan SMK.
 - c) Melakukan kategorisasi status sekolah yaitu sekolah negeri dan swasta.
 - d) Melakukan pelabelan pada peubah asal provinsi, jenis sekolah, status sekolah, dan kategori prestasi akademik setiap sekolah.
2. Membuat model atau pohon keputusan menggunakan data tahun 2019 dengan metode CART.
 - a) Menerapkan *k-fold cross validation* dengan mencobakan jumlah *fold* sebanyak 5 dan 10.
 - b) Menginterpretasikan pohon keputusan yang didapat.
3. Menerapkan model CART pada data tahun 2020.
4. Jika langkah 3 menghasilkan akurasi yang baik, maka model CART akan diterapkan pada data tahun 2021.
5. Jika langkah 3 menghasilkan akurasi yang kurang baik, maka akan dilakukan penggerombolan data tahun 2020.
 - a) Mencobakan beberapa C atau banyak gerombol dengan C yang dicobakan yaitu sebanyak 2, 3, 4, dan 5.
 - b) Memilih C optimal yaitu dengan melihat C dengan rasio S_w/S_b terkecil.
 - c) Melihat karakteristik setiap gerombol.
6. Membuat model baru dengan metode CART menggunakan data hasil penggerombolan tahun 2020.
 - a) Menerapkan *k-fold cross validation* dengan mencobakan jumlah *fold* sebanyak 5 dan 10.
 - b) Menginterpretasikan pohon keputusan yang didapat.
7. Menerapkan model CART tahun 2020 untuk data tahun 2021.
8. Menarik kesimpulan.

3. Hasil dan Pembahasan

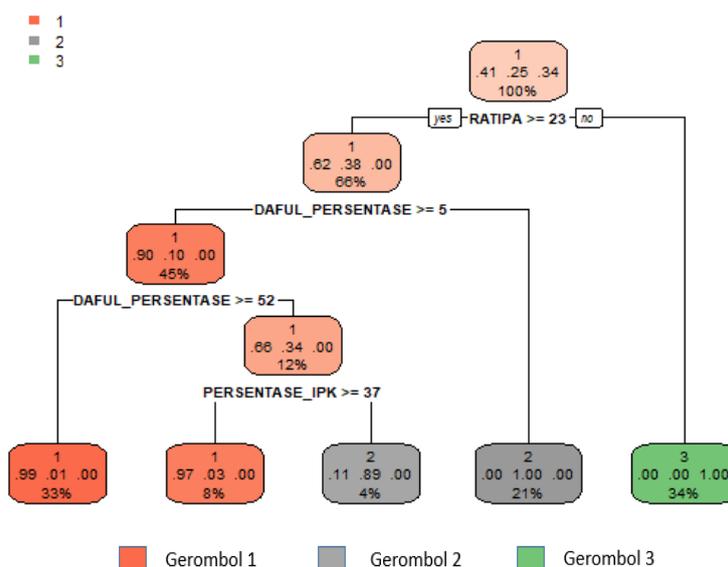
3.1 Model CART Tahun 2019 dan Penerapannya pada Data Tahun 2020

Pada penelitian yang dilakukan oleh Sahili (2021), telah dilakukan pengelompokan sekolah-sekolah yang mendaftar USMI pada minimal satu tahun di tahun 2017, 2018, dan 2019. Berdasarkan pengelompokan tersebut, akan dibuat model klasifikasi yang mengacu pada pengelompokan data pada tahun terbaru yaitu tahun 2019. Akurasi dari kedua jumlah *fold* yang dicobakan pada proses *k-fold cross validation* menghasilkan akurasi yang sama. Akurasi yang didapatkan tergolong baik yaitu mencapai 98,52%. Rata-rata akurasi dari setiap *fold* disajikan pada Tabel 2.

Tabel 2 Akurasi model tahun 2019

Jumlah fold	Akurasi	Kappa	Rataan Sensitivitas	Rataan Spesifitas
5	0,9852	0,9774	0,9840	0,9925
10	0,9852	0,9774	0,9840	0,9925

Pohon keputusan untuk data tahun 2019 dapat dilihat pada Gambar 1. Berdasarkan Gambar 1, didapatkan pemilah awal pada model tersebut adalah peubah rataan nilai IPA pada setiap sekolah. Hal ini menunjukkan bahwa rataan nilai IPA siswa mempengaruhi gerombol setiap sekolah. Jika rataan nilai IPA sekolah tersebut kurang dari 23 maka sekolah tersebut dikategorikan dalam gerombol 3. Peubah lain yang menjadi pemilah pada klasifikasi diatas yaitu peubah persentase mahasiswa yang melakukan daftar ulang dan peubah persentase mahasiswa dengan IPK lebih dari tiga dari setiap sekolah. Simpul terminal yang didapatkan terdiri dari dua simpul gerombol 1, dua simpul gerombol 2, dan satu simpul gerombol 3.

**Gambar 1 Model klasifikasi data tahun 2019**

Dari pohon keputusan pada Gambar 1, dapat dilihat juga karakteristik dari setiap gerombol. Gerombol 1 berisi sekolah- sekolah yang memiliki rataan nilai IPA yang lebih dari 23, persentase daftar ulang yang lebih dari 5, dan persentase IPK yang lebih dari 37. Gerombol 2 berisi sekolah-sekolah yang memiliki rataan nilai IPA yang lebih dari 23, persentase daftar ulang yang kurang dari 5, dan persentase IPK yang kurang dari 37. Sementara itu, gerombol 3 berisi sekolah yang memiliki nilai rataan IPA yang lebih kecil dari 23.

Model yang didapatkan kemudian diaplikasikan pada data tahun 2020 dengan tujuan melihat apakah model tahun 2019 dapat digunakan untuk memprediksi gerombol pada data sekolah tahun 2020. Hasil evaluasi model tahun 2019 pada data tahun 2020 disajikan pada Tabel 3.

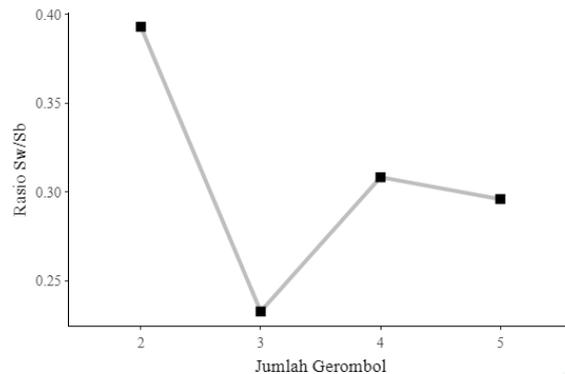
Tabel 3 Akurasi model tahun 2019 dengan data tahun 2020

Statistik	Nilai
Accuracy	0,5722
Kappa	0,3334
Mean Sensitivity	0,5308
Mean Specificity	0,7797

Berdasarkan Tabel 3, hasil evaluasi model tahun 2019 dengan data tahun 2020 menunjukkan masih belum cukup bagus dalam memprediksi karena akurasi yang didapat sebesar 57,22%. Hal ini dapat terjadi karena sekolah-sekolah yang mendaftar pada tahun 2020 memiliki perbedaan dengan sekolah-sekolah yang mendaftar pada tahun 2019. Oleh karena itu, perlu dilakukan penggerombolan ulang untuk data tahun 2020.

3.2 Penggerombolan Data Sekolah Pelamar USMI Tahun 2020

Proses penggerombolan data tahun 2020 dilakukan menggunakan data sekolah yang memiliki minimal satu pelamar pada seleksi USMI tahun 2020 yakni berjumlah 1.184 sekolah. Sebelum melakukan penggerombolan, perlu ditentukan terlebih dahulu jumlah gerombol optimal. Metode yang digunakan adalah dengan menghitung rasio simpangan baku dalam gerombol (S_w) dengan simpangan baku antar gerombol (S_b). Nilai rasio yang kecil mengindikasikan bahwa keragaman di dalam gerombol kecil atau homogen dan keragaman antar gerombol besar atau heterogen.

**Gambar 2 Rasio S_w/S_b dari setiap jumlah gerombol tahun 2020**

Dari hasil perhitungan rasio S_w/S_b pada Gambar 2, saat jumlah gerombol sama dengan tiga, nilai rasio S_w/S_b mengalami penurunan yang signifikan dan kemudian kembali naik lagi pada jumlah gerombol sama dengan empat. Nilai rasio terkecil adalah pada saat $C=3$, maka jumlah gerombol yang dipakai sama dengan 3. Gerombol 1 berisi 644 sekolah, gerombol 2 berisi 232 sekolah, dan gerombol 3 berisi 308 sekolah. Karakteristik dari masing-masing gerombol adalah sebagai berikut.

a. Gerombol 1

Gerombol 1 didominasi sekolah-sekolah yang memiliki nilai yang tinggi pada persentase mahasiswa dengan IPK lebih dari tiga. Selain itu, gerombol 1 juga berisi sekolah-sekolah yang memiliki rata-rata nilai IPA yang tinggi. Pada

peubah prestasi akademik, gerombol 1 didominasi oleh sekolah yang memiliki kategori A dan terdapat beberapa yang berkategori B, C, dan D.

b. Gerombol 2

Gerombol 2 didominasi oleh sekolah-sekolah yang memiliki nilai tinggi pada persentase daftar ulang dan persentase siswa yang diterima. Gerombol 2 juga berisi sekolah yang memiliki peubah rata-rata nilai IPS. Pada peubah prestasi akademik, gerombol 2 didominasi oleh sekolah yang memiliki kategori A tetapi frekuensinya tidak lebih banyak dari gerombol 1. Gerombol 2 juga tidak berisi sekolah yang memiliki kategori C. Selain itu, gerombol 2 didominasi oleh sekolah yang memiliki tingkat komitmen yang tinggi.

c. Gerombol 3

Gerombol 3 berisi sekolah yang memiliki nilai rata-rata yang bukan tertinggi pada setiap peubah. Gerombol 3 didominasi oleh sekolah yang prestasi akademiknya tidak diketahui atau sekolah yang baru mendaftar USMI pada tahun 2020. Selain itu, gerombol 3 berisi sekolah-sekolah yang memiliki tingkat komitmen rendah.

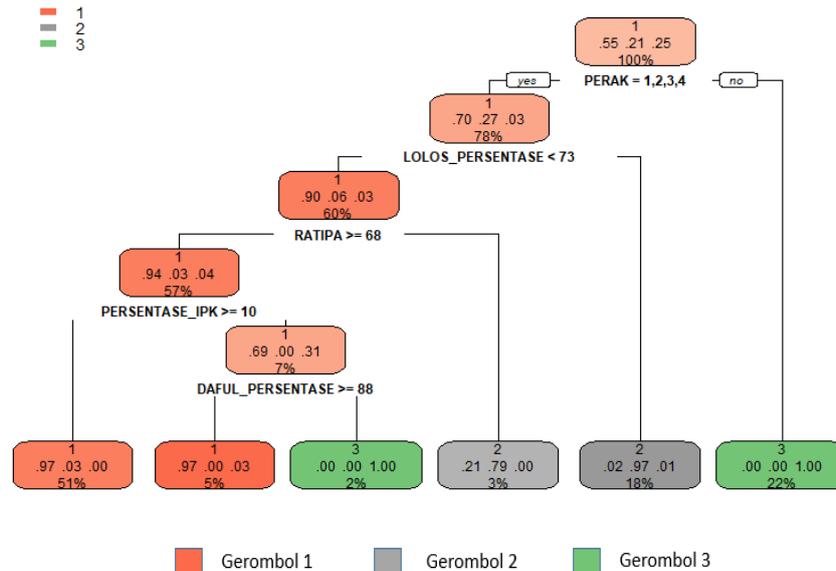
3.3 Model CART Tahun 2020 dan Penerapannya pada Data Tahun 2021

Sama dengan proses pembuatan model pada data tahun 2019, pembentukan model data tahun 2020 juga menggunakan *k-fold cross validation* dan menghasilkan akurasi yang tidak berbeda secara signifikan pada kedua jumlah fold. Akurasi yang didapatkan tergolong baik yaitu mencapai 97,47% pada *10-fold cross validation*. Rata-rata akurasi dari setiap fold disajikan pada Tabel 4.

Tabel 4 Akurasi model tahun 2020

Jumlah fold	Akurasi	Kappa	Rataan Sensitivitas	Rataan Spesifisitas
5	0,9721	0,9534	0,9698	0,9847
10	0,9747	0,9577	0,9737	0,9863

Pohon keputusan tahun 2020 dapat dilihat pada Gambar 3. Model pohon keputusan tahun 2020 terlihat lebih besar dari model data tahun 2019. Pemilah awal pada model pohon keputusan di atas merupakan peubah prestasi akademik, artinya kategori prestasi akademik dari suatu sekolah sangat penting dalam proses pengklasifikasian dimana jika suatu sekolah tidak diketahui kategori prestasi akademiknya maka akan diklasifikasikan ke dalam gerombol 3. Peubah lain yang juga menjadi pemilah pada klasifikasi ini adalah persentase siswa lolos seleksi, rata-rata nilai IPA, persentase mahasiswa dengan IPK lebih dari tiga, dan persentase siswa yang melakukan daftar ulang. Simpul terminal pada model tersebut terdiri dari dua simpul gerombol 1, dua simpul gerombol 2, dan dua simpul gerombol 3.



Gambar 3 Model klasifikasi tahun 2020

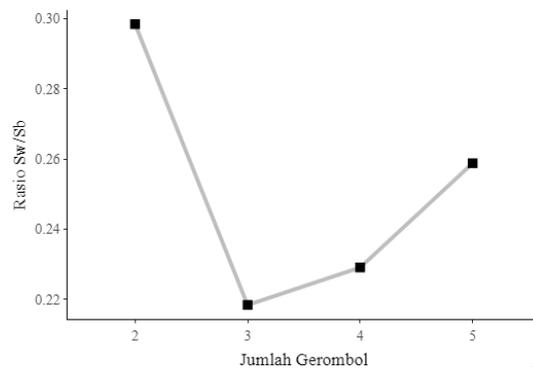
Setelah didapatkan model pada Gambar 3, model tersebut akan diterapkan pada data tahun 2021. Proses ini dilakukan untuk melihat apakah model yang sudah dibentuk dapat memperkirakan gerombol dari suatu sekolah untuk tahun selanjutnya. Akurasi dari evaluasi model klasifikasi tahun 2020 pada data tahun 2021 disajikan pada Tabel 5.

Tabel 5 Akurasi proses evaluasi model tahun 2020 dengan data tahun 2021

Statistik	Nilai
Accuracy	0.4434
Kappa	0.1719
Mean Sensitivity	0.4589
Mean Specificity	0.7313

Berdasarkan Tabel 5, hasil evaluasi model tahun 2020 dengan data tahun 2021 menunjukkan masih belum cukup bagus dalam memprediksi karena akurasi yang didapat sebesar 44,34%. Hal ini dapat terjadi karena sekolah-sekolah yang mendaftar pada tahun 2021 memiliki perbedaan dengan sekolah-sekolah yang mendaftar pada tahun 2020. Oleh karena itu, perlu dilakukan penggerombolan ulang untuk data tahun 2021.

3.4 Penggerombolan Data Sekolah Pelamar USMI Tahun 2021



Gambar 4 Rasio Sw/Sb dari setiap gerombol tahun 2021

Dari hasil perhitungan rasio Sw/Sb pada Gambar 4, didapatkan $C=3$ merupakan jumlah gerombol optimal. Gerombol pertama berisi 442 sekolah, gerombol kedua berisi 370 sekolah, dan gerombol ketiga berisi 389 sekolah. Karakteristik dari masing-masing gerombol adalah sebagai berikut.

a. Gerombol 1

Gerombol 1 berisi sekolah-sekolah yang memiliki nilai yang tinggi pada persentase mahasiswa dengan IPK lebih dari tiga. Selain itu juga gerombol 1 berisi sekolah-sekolah yang memiliki rata-ran nilai IPS yang tinggi. Pada peubah prestasi akademik, gerombol 1 didominasi oleh sekolah yang memiliki kategori D, tetapi jumlah sekolah yang memiliki kategori A, B, dan C pada gerombol 1 lebih banyak dari gerombol lain.

b. Gerombol 2

Gerombol 2 didominasi oleh sekolah-sekolah yang memiliki nilai tinggi pada persentase daftar ulang dan persentase siswa yang diterima. Pada peubah prestasi akademik, gerombol 2 didominasi oleh sekolah yang memiliki kategori D dan juga hanya sedikit sekolah yang memiliki kategori A, B, C, dan D pada gerombol 2.

c. Gerombol 3

Gerombol 3 berisi sekolah yang memiliki nilai rata-ran nilai IPA yang tinggi. Pada peubah prestasi akademik, gerombol 3 didominasi oleh sekolah yang prestasi akademiknya tidak diketahui atau sekolah yang baru mendaftar USMI pada tahun 2021. Selain itu, gerombol 3 berisi sekolah-sekolah yang memiliki tingkat komitmen rendah.

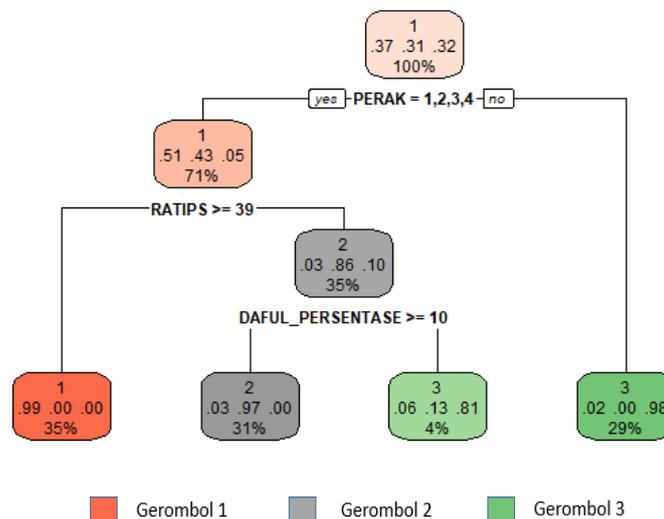
3.5 Model Klasifikasi CART Data Tahun 2021

Sama dengan proses pembuatan model pada data tahun 2020, pembentukan model data tahun 2021 juga menggunakan k-fold cross validation dan menghasilkan akurasi yang tidak berbeda secara signifikan pada kedua jumlah fold. Akurasi yang didapatkan tergolong baik yaitu mencapai 98% pada 10-fold cross validation. Rata-rata akurasi dari setiap fold disajikan pada Tabel 6.

Tabel 6 Akurasi model tahun 2021

Jumlah fold	Akurasi	Kappa	Rataan Sensitivitas	Rataan Spesifisitas
5	0,9775	0,9662	0,9780	0,9887
10	0,9800	0.9699	0,9798	0,9899

Model pohon keputusan tahun 2021 disajikan pada Gambar 5. Pada Gambar 5 terlihat bahwa pemilah awal pada pohon klasifikasi diatas adalah peubah prestasi akademik, yang berarti peubah prestasi akademik sangat berpengaruh pada pengelompokkan sekolah tahun 2021. Peubah lain yang berpengaruh juga adalah rataan nilai IPS dan persentase siswa daftar ulang. Simpul terminal dari pohon keputusan tersebut terdiri dari satu simpul gerombol 1, satu simpul gerombol 2, dan dua simpul gerombol 3.



Gambar 5 Model klasifikasi tahun 2021

4. Simpulan

Pemodelan klasifikasi data tahun 2019 menghasilkan model dengan akurasi yang cukup baik. Peubah yang berpengaruh pada klasifikasi tahun 2019 adalah rataan nilai IPA, persentase siswa daftar ulang, dan persentase mahasiswa dengan IPK lebih dari tiga. Namun, evaluasi model tahun 2019 dengan data tahun 2020 masih belum cukup memprediksi karena akurasi yang didapat sebesar 55,22%. Oleh karena itu, dilakukan penggerombolan terhadap data tahun 2020. Gerombol yang dihasilkan sebanyak tiga gerombol dengan karakteristik gerombol 1 berisi sekolah yang memiliki nilai yang tinggi pada peubah persentase mahasiswa dengan IPK lebih dari tiga dan peubah rataan nilai IPA, gerombol 2 berisi sekolah yang memiliki nilai yang tinggi pada peubah persentase siswa diterima, persentase siswa daftar ulang, dan rataan nilai IPS. Sementara itu, gerombol 3 berisi sekolah dengan nilai yang rendah pada setiap peubah dan berisi sekolah-sekolah yang baru mendaftar.

Hasil gerombol tahun 2020 dipakai untuk membuat model klasifikasi. Model yang didapatkan memiliki akurasi yang cukup baik dengan peubah yang berpengaruh yaitu prestasi akademik, persentase siswa diterima, rata-rata nilai IPA, persentase IPK, dan persentase daftar ulang. Namun, evaluasi model tahun 2020 dengan data tahun 2021 masih belum cukup memprediksi karena akurasi yang didapat sebesar 44,34%. Oleh karena itu, untuk melihat pengelompokan sekolah pelamar USMI, perlu dilakukan penggerombolan sekolah pada setiap tahun berjalan. Pohon klasifikasi yang didapatkan tidak dapat digunakan untuk memprediksi data penerimaan mahasiswa baru, melainkan hanya dapat digunakan untuk eksplorasi gerombol pada tahun berjalan. Peubah yang berpengaruh pada setiap tahunnya adalah peubah persentase siswa daftar ulang.

Daftar Pustaka

- Alverina D, Chrismanto AR, Santosa RG. 2018. Perbandingan Akurasi Algoritma C4.5 dan CART dalam Memprediksi Kategori Indeks Prestasi Mahasiswa. *Jurnal Teknologi dan Sistem Komputer*. 6(2): 76 – 83. [Diunduh 17 Feb 2022]. DOI: 10.14710/jtsiskom.6.2.2018.76-83.
- Aritonang NE, Rusgiyono A, Rita R. 2016. Klasifikasi status kerja pada angkatan kerja Kota Semarang tahun 2014 menggunakan metode CHAID dan CART. *Jurnal Gaussian*. 5(1):183–192. [Diunduh 27 Jul 2022].
- Bunkers MJ, Miller JR, DeGaetano AT. 1996. Definition of climate region in the northern plains using an objective cluster modification technique. *Journal of Climate*. 9(1): 130-146. [Diunduh 11 Jul 2022]. DOI: 10.1175/15200442(1996)009%3C0130:DOCRIT%3E2.0.CO;2.
- Breiman L, Friedman JH, Olshen RA, Stone CJ. 1993. *Classification and Regression Trees*. New York (US): Chapman and Hall.
- Diego IMD, Redondo AR, Fernandez RR, Navarro J, Moguerza JM. 2021. General Performance Score for Classification Problems. *Applied Intelligence*. 52: 12049 – 12063. [Diunduh 8 Jul 2022]. DOI: 10.1007/s10489-021-03041-7.
- Grandini M, Bagli E, Visani G. 2020. Metrics for multi-class classification: an overview. arXiv preprint. [Diunduh 4 Jul 2022].
- Han J, Kamber M, Pei J. 2012. *Data Mining Concepts and Techniques*. 3rd Edition. Waltham (US): Elsevier Inc.
- Huang Z. 1998. Extension to the K-Means algorithm for clustering large data sets with categorical values. *Data Mining and Knowledge Discovery*. 2(3): 283– 304. [Diunduh 18 Jul 2022].
- Komalasari WB. 2007. Metode Pohon Regresi Untuk Eksploratori Data Dengan Peubah yang Banyak dan Kompleks. *Informatika Pertanian*. 16(1) :967 – 980. [Diunduh 26 Nov 2021].
- James G, Witten D, Hastie T, Tibshirani R, 2013. *An Introduction to Statistical Learning* (Vol. 112, p. 18). New York (US): Springer.
- Kader GD, Perry M. 2007. Variability for categorical variables. *Journal of Statistics Educations*. 15(2): 1–16.
- Lewis RJ. 2000. An introduction to classification and regression tree (CART) analysis. Annual Meeting of the Society for Academic Emergency Medicine. California (USA): UCLA Medical Center.
- Manliguez C. 2016. Generalized Confusion Matrix for Multiple Classes. Researchgate.[Diunduh 8 jul 2022].
- Okada T. 1999. Sum of squares decomposition for categorical data. *Kwansei Gakuin Studies in Computer Science*. 14: 1–6.

- Pratiwi R. 2020. Perbandingan Klasifikasi Algoritma C5.0 Dan Classification And Regression Tree (Studi Kasus : Data Sosial Kepala Keluarga Masyarakat Desa Teluk Baru Kecamatan Muara Ancalong Tahun 2019) [Skripsi]. Samarinda : Universitas Mulawarman.
- Sahili SU.2021. Penggerombolan Sekolah Dalam Penerimaan Mahasiswa Baru Sekolah Vokasi Ipb Jalur Usmi Menggunakan Algoritme *K-Prototypes* [Skripsi]. Bogor : Institut Pertanian Bogor.
- Sinambela YES.2008. Penerapan Metode Pohon Klasifikasi dengan Algoritma Cart pada Data Status Daerah Kabupaten di Indonesia [Skripsi]. Bogor : Institut Pertanian Bogor.
- Sulastri S. 2021. Penggerombolan Perusahaan Budidaya Ikan di Indonesia dengan Metode K-Prototypes, Two Step Cluster (Tsc), dan Gower [Skripsi]. Bogor : Institut Pertanian Bogor.