

Identifikasi Faktor-Faktor yang Memengaruhi Prestasi Mahasiswa Program Sarjana di Institut Pertanian Bogor Menggunakan Metode Chaid*

Ragsa Endahas A¹, Akbar Rizki^{2‡}, Mohammad Masjkur³

^{1,2,3}Department of Statistics, IPB University, Indonesia

[‡]corresponding author: akbar.ritzki@apps.ipb.ac.id

Copyright © 2022 Ragsa Endahas A, Akbar Rizki, and Mohammad Masjkur. This is an open-access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

IPB University (IPB) is one of the best universities in Indonesia, based on the Ministry of Education and Culture (Kemendikbud) clustering in 2020. As the best university, IPB requires efforts to improve the quality of its education. One of these efforts is to improve student achievement. This study aims to identify the factors that influence the competition and non-competition achievements of undergraduate students at IPB. The data used are achievement data (academic year 2016/2017 to 2020/2021) from the Directorate of Student Affairs and Career Development (Ditmawa) of IPB and demographic data of undergraduate level IPB students (entry year 2016/2017 to 2019/2020) from the Directorate of Administration and Education (Dit-Ap) IPB. The analytical method used in this study is the Chi-square Automatic Interaction Detection (CHAID) classification method. There was an imbalance of data on the Student Achievement response variable. Therefore, in this study, unbalanced data handling was also carried out by resampling in the form of oversampling, undersampling, and over-undersampling methods. The results showed that the classification using CHAID analysis with resampling in the form of oversampling with a balance accuracy of 73.7% resulted in the best classification performance. The factors that influence student achievement are 11 variables, and the 3 most influential variables are variables of year of admission, department, and last GPA.

Keywords: achievement; CHAID; main performance indicators of universities; university; university classification.

* Received: Nov 2021; Reviewed: May 2022; Published: May 2022

1. Pendahuluan

Indikator Kinerja Utama (IKU) Perguruan Tinggi Negeri adalah sebuah ukuran atau indikator kerja instansi perguruan tinggi negeri yang diciptakan oleh lembaga pemerintahan yaitu Direktorat Jenderal Pendidikan Tinggi (DIKTI) (DIKTI, 2020). Harapannya dengan adanya IKU, setiap Perguruan Tinggi Negeri (PTN) mempunyai acuan yang jelas dalam meningkatkan kualitas pendidikannya. DIKTI juga memberikan reward dana bagi Perguruan Tinggi (PT) yang mampu mencapai standar nilai dari indikator perguruan tinggi negeri ini. IKU merupakan salah satu upaya pemerintah dalam meningkatkan kualitas pendidikan perguruan tinggi negeri di Indonesia. Selain IKU, juga terdapat program Merdeka Belajar (MB). Menurut Kemendikbud (2020), Program ini berupaya meningkatkan kualitas pendidikan secara langsung yaitu melalui peningkatan kompetensi lulusan, baik soft skills maupun hard skills di kalangan mahasiswa/peserta didik perguruan tinggi yang nantinya menjadi nilai tambah dalam meningkatkan daya saing untuk mendapat pekerjaan. Kemendikbud juga berupaya merumuskan kualitas dari 2.136 perguruan tinggi yang terdata di PD Dikti melalui program Klasterisasi Kemendikbud (KK). Adanya klasterisasi diharapkan bisa memilah perguruan tinggi untuk kepentingan pembinaan perguruan tinggi sehingga mampu memberikan porsi yang pas antara setiap tingkatan klasifikasinya. Ketiga hal tersebut, mempunyai tujuan yang sama yaitu upaya dalam meningkatkan kualitas pendidikan di perguruan tinggi khususnya perguruan tinggi negeri di Indonesia. Peubah Prestasi Mahasiswa (PM) adalah salah satu peubah yang diukur pada IKU, MB, dan KK. Menurut Roulin dan Bangerter 2013, prestasi mahasiswa juga akan meningkatkan daya saing mahasiswa untuk mendapatkan pekerjaan. Hal ini menunjukkan bahwa peubah ini penting untuk diperhatikan. Apabila suatu perguruan tinggi memiliki PM yang tinggi, maka perguruan tinggi akan memiliki peluang besar untuk memiliki ketercapaian IKU dan program MB yang tinggi serta termasuk ke dalam klasterisasi PT yang baik menurut Kemendikbud.

IPB adalah salah satu perguruan tinggi terbaik di Indonesia, yang ditunjukkan oleh hasil pemeringkatan perguruan tinggi oleh Kemendikbud. Pada tahun 2019 dan Tahun 2020 IPB berturut-turut berhasil mendapatkan peringkat tiga dan peringkat satu berdasarkan klasterisasi PT kemendikbud. Oleh karena itu, untuk mempertahankan prestasi tersebut sekaligus meningkatkan kualitas pendidikannya, IPB membutuhkan upaya-upaya tertentu. Salah satunya adalah upaya dalam hal meningkatkan prestasi mahasiswa baik prestasi lomba maupun non lomba. IPB memiliki banyak mahasiswa, dimana diantara mahasiswa tersebut, ada yang memiliki prestasi dan tidak memiliki prestasi. Faktor-faktor yang berpengaruh terhadap prestasi mahasiswa diantaranya adalah faktor dari diri mahasiswa sendiri (internal) dan faktor dari lingkungan mahasiswa (eksternal). Mahasiswa program sarjana merupakan populasi mahasiswa terbesar yaitu sebesar 59.7% dari total mahasiswa IPB (IPB, 2019). Oleh karena itu, penelitian ini akan berfokus pada faktor-faktor yang berpengaruh terhadap prestasi mahasiswa program sarjana IPB. Salah satu metode yang dapat digunakan untuk mengidentifikasi faktor-faktor tersebut adalah metode klasifikasi.

Salah satu metode klasifikasi yang dapat digunakan adalah metode Chi-square Automatic Interaction Detection (CHAID) (Pertiwi, 2013). Menurut Díaz-Pérez 2020, CHAID merupakan salah satu klasifikasi non parametrik yang paling efektif dan paling sering digunakan, sebab metode CHAID bisa menggunakan banyak peubah untuk klasifikasi sekaligus membantu mengidentifikasi faktor-faktor dari peubah respon. Menurut Kim H dan Loh W 2001, pohon klasifikasi CHAID menggunakan *nonbinary*

split atau *multiway split* yang berakibat memudahkan interpretasi serta mengurangi kebiasaan dalam interpretasi karakteristik. Selain itu, CHAID dianggap metode yang paling adaptif dalam menangani data ordinal maupun nominal, bila ada data yang hilang pun atau missing value, CHAID dapat menanganinya dengan relatif mudah (Murphy, 2013).

2. Metodologi

2.1 Data

Tabel 1 Rincian Peubah yang digunakan

Nama peubah	Jenis	Keterangan
Prestasi (Y1)	Nominal	1 = Berprestasi, 0 = Tidak Memiliki Prestasi
Departemen (X1)	Nominal	A1 = Manajemen s.d K1 = Bisnis
Fakultas (X2)	Nominal	A = Pertanian s.d K = Sekolah Bisnis
Jenis kelamin (X3)	Nominal	1 = Laki-Laki, 0 = Perempuan
Jalur masuk (X4)	Nominal	1 = SNMPTN, 2 = SBMPTN, 3 = BUD, 4 = UTMI / UTM, 5 = Ketua Osis, 6 = Afirmasi, 7 = Asing, 8 = PIN
Daerah asal (X5)	Nominal	Daerah asal mengacu ke Lembaga Layanan DIKTI (LLDIKTI) tahun 2021 untuk daerah 1-12 dan daerah 13 adalah luar negeri. Daerah 1 = Sumatera Utara dan daerah Istimewa Aceh s.d Daerah 12 = Maluku, Papua, dan Papua Barat, dan Daerah 13 = Luar Negeri
Jenis Sekolah Asal (X6)	Nominal	0 = Swasta, 1 = Negeri
IPK terakhir (X7)	Ordinal	1 = 0-2.88, 2 = 2.88-3.14, 3 = 3.14-3.35, 4 = 3.35-3.57, 5 = 3.57-4
UKM (X8)	Nominal	0 = Tidak Mengikuti UKM, 1 = Mengikuti UKM
Tahun Masuk (X9)	Nominal	1 = Tahun 2016/2017 s.d 4 = Tahun 2019/2020
SKS Kumulatif (X10)	Ordinal	1 = 0-38, 2 = 38-80, 3 = 80-119, 4 = 119-130, 5 = 130-159
Status Kelanjutan Studi (X11)	Nominal	0 = Tanpa Syarat, 1 = Peringatan, 2 = Peringatan Keras, 3 = Dikeluarkan

Data yang digunakan pada penelitian ini merupakan data yang diperoleh dari Direktorat Administrasi dan Pendidikan (Dit-AP) IPB dan Direktorat Kemahasiswaan dan Pengembangan Karir (Ditmawa) IPB. Data terdiri dari data prestasi mahasiswa IPB tingkat sarjana (tahun ajaran 2016/2017 s.d 2020/2021) dan data demografi mahasiswa IPB tingkat sarjana (tahun masuk 2016/2017 s.d 2019/2020) yang selanjutnya disebutkan sebagai populasi mahasiswa. Terdapat 11 peubah penjelas dan 1 peubah respon yang akan digunakan dalam penelitian. Peubah respon yang diamati adalah prestasi lomba dan non-lomba mahasiswa yang selanjutnya akan disebutkan sebagai prestasi mahasiswa. Peubah-peubah yang digunakan dalam penelitian ini dapat dilihat pada Tabel 1. Peubah Indeks Prestasi kumulatif (IPK) dan Satuan Kredit Semester (SKS) Kumulatif mahasiswa dikategorikan secara proporsional berdasarkan jumlah mahasiswa yang berada pada masing-masing kategori. Peubah Daerah asal mengacu ke Lembaga Layanan DIKTI (LLDIKTI) tahun

2021 untuk satu (1) sampai dengan dua belas (12) dengan penambahan daerah tiga belas (13) yaitu luar negeri.

2.2 Metode Penelitian

Berikut adalah penjelasan mengenai tahapan analisis yang digunakan dalam penelitian ini:

1. Melakukan praproses data:
 - a. Menggabungkan data yang didapat dari Dit-Ap dan Ditmawa.
 - b. Melakukan pelabelan data sesuai kategori peubah respon yang diharapkan dalam Tabel 1.
2. Eksplorasi data:
 - a. Melakukan analisis deskriptif terhadap peubah penjelas dan peubah respon.
 - b. Mengidentifikasi peubah-peubah penjelas yang memiliki hubungan dengan peubah respon menggunakan uji kebebasan khi-kuadrat dengan taraf nyata 0.05.
3. Melakukan analisis CHAID pada data tidak seimbang.
4. Melakukan analisis CHAID:
 - a. Pada data yang telah dilakukan resampling dengan metode *Oversampling*.
 - b. Pada data yang telah dilakukan resampling dengan metode *Undersampling*.
 - c. Pada data yang telah dilakukan resampling dengan metode *Over-Undersampling*.
5. Mengevaluasi model/kinerja klasifikasi dan pemilihan model terbaik pada tahap empat (4), dengan melihat nilai *balance accuracy* dan *Cross Validation*.
6. Menginterpretasi hasil dengan melihat peubah-peubah yang memengaruhi prestasi mahasiswa.

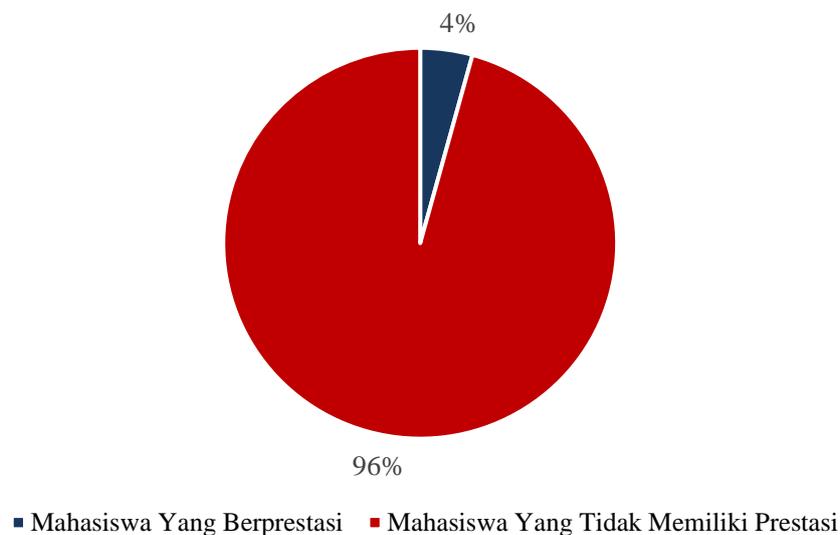
3. Hasil dan Pembahasan

3.1 Eksplorasi Data

3.1.1. Gambaran Umum Data

Data yang digunakan dalam penelitian ini merupakan data Direktorat Administrasi dan Pendidikan (Dit-Ap) IPB dengan objek prestasi mahasiswa IPB tingkat sarjana (tahun ajaran 2016/2017 s.d 2020/2021) dan Direktorat Kemahasiswaan dan Pengembangan Karir (Ditmawa) IPB dengan objek data demografi mahasiswa IPB tingkat sarjana (tahun masuk 2016/2017 s.d 2019/2020). Data tersebut sebanyak 13515 amatan (mahasiswa) yang terdiri atas 579 mahasiswa (4%) berstatus memiliki prestasi dan 12936 mahasiswa (96%) berstatus tidak mempunyai prestasi seperti yang ditunjukkan pada Gambar 1. Prestasi yang dimaksud adalah prestasi lomba maupun

non-lomba dengan jenis kepesertaan individu maupun tim yang berskala nasional maupun internasional.



Gambar 1 Persentase mahasiswa berdasarkan status kepemilikan prestasi

3.1.2. Nilai Khi-kuadrat

Tabel 2 Khi kuadrat peubah penjelas mahasiswa yang berprestasi

Peubah	Nilai khi-kuadrat	Nilai-p
Departemen (X1)	113.284	.000*
Fakultas (X2)	24.151	.004*
Jenis kelamin (X3)	21.838	.000*
Jalur masuk (X4)	18.315	.011*
Daerah asal (X5)	24.104	.020*
Jenis Sekolah (X6)	4.094	.043*
IPK terakhir (X7)	60.344	.000*
UKM (X8)	23.586	.000*
Tahun masuk (X9)	148.345	.000*
SKS Kumulatif (X10)	58.449	.000*
Status Kelanjutan Studi (X11)	1.706	.636

*Signifikan pada taraf nyata 0,05

Tabel 2 menyajikan informasi hasil uji kebebasan khi-kuadrat antara peubah respon dan masing-masing peubah penjelas. Terdapat sepuluh dari sebelas peubah yang memiliki hubungan signifikan atau memiliki nilai p-value lebih kecil dari taraf nyata 0.05 terhadap peubah respon atau mahasiswa berprestasi yaitu peubah departemen, fakultas, jenis kelamin, jalur masuk, daerah asal, jenis sekolah, IPK terakhir, UKM, tahun masuk dan SKS kumulatif. Karena memiliki nilai p-value lebih besar daripada 0.05, maka Peubah Status Kelanjutan Studi tidak berpengaruh terhadap prestasi mahasiswa. Peubah Status Kelanjutan Studi merupakan informasi status akademik mahasiswa, yang terdiri dari tanpa peringatan, peringatan, peringatan keras,

dikeluarkan.

3.2 Analisis Chaid pada Data Tidak Seimbang dan Penanganan Data Tidak Seimbang

Data yang digunakan dalam penelitian ini menunjukkan data tidak seimbang. Hal ini dikarenakan dari sebanyak 13515 mahasiswa, mahasiswa yang berprestasi hanya terdiri dari 579 (4%) mahasiswa saja, sedangkan yang tidak mempunyai prestasi terdiri dari 12936 (96%) mahasiswa. Hasil analisis CHAID pada data tidak seimbang menunjukkan bahwa terdapat sepuluh peubah penjelas yang berasosiasi dengan peubah respon (prestasi mahasiswa). Peubah penjelas yang berpengaruh tersebut adalah daerah asal, UKM, departemen, IPK, jenis kelamin, tahun masuk, SKS kumulatif, jenis sekolah, fakultas, dan jalur masuk.

Kinerja klasifikasi CHAID pada data tidak seimbang dapat dilihat pada Tabel 3. Tabel tersebut menunjukkan bahwa ketepatan klasifikasi yang dihasilkan dengan menggunakan data tidak seimbang memperoleh akurasi pada kelas mahasiswa yang berprestasi sebesar 0%, akurasi untuk kelas mahasiswa yang tidak berprestasi sebesar 100% dan *balance accuracy* 50%. Walaupun mendapat ketepatan klasifikasi yang besar, data tidak seimbang dikhawatirkan akan mengurangi ketepatan klasifikasi model. Menurut Spelman dan Porkodi (2018), melakukan klasifikasi pada data yang tidak seimbang akan mengakibatkan menurunnya ketepatan klasifikasi karena klasifikasi dalam memprediksi kelas minoritas menjadi tidak akurat. Oleh karena itu, data tidak seimbang perlu ditangani. Menangani data tidak seimbang bisa menggunakan metode resampling yaitu menyeimbangkan data kelas minoritas dan mayoritas agar kedua kelas seimbang dengan eliminasi, duplikasi maupun gabungannya (Hustini 2020). Penelitian ini akan menggunakan resampling yang berupa *Oversampling*, *Undersampling*, dan *Over-undersampling* untuk menyeimbangkan data, yang selanjutnya akan dipakai kembali untuk analisis CHAID.

Tabel 3 Ketepatan klasifikasi CHAID

Aktual	Prediksi		Persentase benar
	Tidak Berprestasi	Berprestasi	
Tidak Berprestasi	12936	0	100%
Berprestasi	579	0	0%
<i>Balance Accuracy</i>			50%

3.3 Analisis CHAID pada Data yang Diseimbangkan

3.3.1. Analisis CHAID pada Data Hasil Metode Oversampling

Resampling data dengan metode *oversampling* yaitu menambahkan data pada kelas minoritas dalam menyeimbangkan data, sehingga menghasilkan 25711 amatan (Mahasiswa) yang terdiri dari 12775 (49.69%) Mahasiswa berprestasi dan 12936

(50.31%) Mahasiswa yang tidak berprestasi. Analisis CHAID dengan *oversampling* menghasilkan sebelas peubah penjelas yang berasosiasi terhadap peubah respon mahasiswa yang berprestasi. Ketepatan klasifikasi yang ditunjukkan pada tabel 4, *balance accuracy*, spesifisitas, dan sensitivitas yang dihasilkan oleh analisis CHAID dengan *oversampling* adalah sebesar 73.7%, 66.1%, dan 81.4%.

Tabel 4 Ketepatan Klasifikasi CHAID dengan *Oversampling*

Aktual	Prediksi		Persentase benar
	Tidak Berprestasi	Berprestasi	
Tidak Berprestasi	8556	4380	66.1%
Berprestasi	2374	10401	81.4%
<i>Balance Accuracy</i>			73.7%

3.3.2. Analisis CHAID pada Data Hasil Metode *Undersampling*

Resampling data dengan metode *undersampling* yaitu pengurangan data pada kelas mayoritas dalam menyeimbangkan data, sehingga menghasilkan 1152 amatan (Mahasiswa) yang terdiri dari 579 (50.26%) Mahasiswa berprestasi dan 573 (49.74%) Mahasiswa yang tidak berprestasi. Analisis CHAID dengan *undersampling* ketepatan klasifikasi yang ditunjukkan pada tabel 5, *balance accuracy*, spesifisitas, dan sensitivitas yang dihasilkan oleh analisis CHAID dengan *undersampling* adalah sebesar 68.1%, 54.8%, dan 81.2%.

Tabel 5 Ketepatan klasifikasi CHAID dengan *Undersampling*

Aktual	Prediksi		Persentase benar
	Tidak Berprestasi	Berprestasi	
Tidak Berprestasi	314	259	54.8%
Berprestasi	109	470	81.2%
<i>Balance Accuracy</i>			68.1%

3.3.3. Analisis CHAID pada Data Hasil Metode *Over-undersampling*

Resampling data dengan metode *over-undersampling* yaitu menambahkan data pada kelas minoritas dan mengurangi data pada kelas mayoritas dalam menyeimbangkan data, sehingga menghasilkan 13515 amatan (Mahasiswa) yang terdiri dari 6671

(49.36%) Mahasiswa berprestasi dan 6844 (50.64%) Mahasiswa yang tidak berprestasi. Analisis CHAID dengan *over-undersampling* menghasilkan ketepatan klasifikasi yang ditunjukkan pada tabel 6, *balance accuracy*, spesifisitas, dan sensitivitas yang dihasilkan oleh analisis CHAID dengan *over-undersampling* adalah sebesar 73%, 69.7%, dan 76.4%.

Tabel 6 Ketepatan klasifikasi CHAID dengan *over-sampling*

Aktual	Prediksi		Persentase benar
	Tidak Berprestasi	Berprestasi	
Tidak Berprestasi	4769	2075	69.7%
Berprestasi	1572	5099	76.4%
<i>Balance Accuracy</i>			73%

4. Memilih Metode Terbaik

4.1 Memilih Metode Terbaik

Tabel 7 menampilkan komposisi data pada peubah respon status mahasiswa berprestasi sebelum dan sesudah diterapkan *resampling*. Perlakuan data dengan *resampling* menghasilkan jumlah komposisi data yang berbeda-beda. Tetapi, komposisi data yang dihasilkan *resampling* mempunyai proporsi mahasiswa berprestasi dan mahasiswa tidak berprestasi yang lebih seimbang.

Tabel 7 komposisi data pada peubah respon mahasiswa berprestasi sebelum dan sesudah diterapkan *resampling*

Kategori respon	Data tidak seimbang	<i>Oversampling</i>	<i>Undersampling</i>	<i>Over-undersampling</i>
Mahasiswa Berprestasi	579 (4%)	12775 (49.69%)	579 (50.26%)	6671 (49.36%)
Mahasiswa Tidak Berprestasi	12936 (96%)	12936 (50.31%)	573 (49.74%)	6844 (50.64%)
Total	13115 (100%)	25711 (100%)	1152 (100%)	13515 (100%)

Tabel 8 menyajikan ketepatan klasifikasi pada masing-masing model. Besarnya nilai *balance accuracy*, perbedaan nilai spesifisitas dan sensitivitas dari model, maka

model yang akan digunakan untuk mengklasifikasi mahasiswa yang berprestasi pada penelitian ini adalah Model hasil analisis CHAID dengan penyeimbangan data menggunakan metode *oversampling* yang selanjutnya disebutkan sebagai model terbaik. Penanganan data tidak seimbang pada Model terbaik mampu meningkatkan *balance accuracy* model sebesar 23.7% serta dapat meningkatkan performa model dalam memprediksi kelas minoritas. Hal ini dapat terlihat dari peningkatan nilai sensitivitas model sebesar 81.4%. Nilai spesifisitas model mengalami penurunan, namun penurunan nilainya lebih kecil dibandingkan dengan peningkatan nilai sensitivitas, yaitu menjadi 66.1%.

Tabel 8 Ketepatan klasifikasi pada peubah respon mahasiswa berprestasi masing-masing model

Kategori respon	Data tidak seimbang	<i>Oversampling</i>	<i>Undersampling</i>	<i>Over-undersampling</i>
Sensitivitas	0%	81.4%	81.2%	76.4%
Spesifisitas	100%	66.1%	54.8%	69.7%
<i>Balance Accuracy</i>	50%	73.7%	68.1%	73%

4.2 Pohon Klasifikasi CHAID Pada Metode Terbaik

Pemecah simpul utama pada diagram pohon yang dihasilkan pada model terbaik adalah peubah tahun masuk. Mahasiswa dengan tahun masuk 2016/2017 memiliki kecenderungan untuk berprestasi paling tinggi, sebaliknya mahasiswa dengan tahun masuk 2019/2020 memiliki kecenderungan untuk berprestasi paling rendah. Selanjutnya, mahasiswa berasal dari departemen Ilmu Ekonomi Syariah dan Manajemen Sumberdaya Lahan memiliki potensi yang tinggi untuk berprestasi. Mahasiswa dengan nilai IPK terakhir yang diantara 2.88 sampai dengan 4 berkontribusi positif atau berpotensi tinggi terhadap mahasiswa berprestasi, dengan mahasiswa yang memiliki nilai IPK terakhir dibawah 2.88 berkontribusi negatif atau berpotensi rendah terhadap mahasiswa berprestasi. Mahasiswa yang mengikuti UKM atau jalur masuk SNMPTN memiliki berkontribusi positif atau berpotensi tinggi terhadap mahasiswa berprestasi serta mahasiswa semakin banyak SKS Kumulatif yang telah dijalani mahasiswa, maka mahasiswa tersebut berpeluang lebih tinggi untuk berprestasi.

5. Simpulan

Prestasi mahasiswa di IPB dipengaruhi oleh seluruh peubah yang digunakan pada penelitian ini, tiga peubah yang memiliki hubungan paling erat dengan prestasi mahasiswa adalah Tahun Masuk, Departemen, dan IPK terakhir. Model hasil analisis CHAID dengan penyeimbangan data menggunakan metode *oversampling*, mampu meningkatkan ketepatan klasifikasi pada data yang tidak seimbang dengan nilai *balance accuracy*, spesifitas dan sensitivitas secara berturut-turut sebesar 73.7%, 66.1% dan 81.4%. Mahasiswa yang masuk ke IPB melalui jalur masuk SNMPTN,

berasal dari departemen Ilmu Ekonomi Syariah atau Manajemen Sumberdaya Lahan, memiliki IPK terakhir lebih besar dari 2.88, dan mengikuti UKM memiliki potensi besar untuk berprestasi dibandingkan dengan yang lainnya.

Daftar Pustaka

- Agresti A. 1990. *Categorical Data Analysis*. USA: University of Florida.
- Breiman L, JH Friedman, RA Olshen, CJ Stone. 1993. *Classification and Regression Trees*. New York: Chapman and Hall.
- Díaz-Pérez F M, García-González C G, Fyall F. 2020. The use of the CHAID algorithm for determining tourism segmentation: A purposeful outcome, *Heliyon*, Volume 6, Issue 7
- Direktorat Jendral Pendidikan Tinggi [DIKTI]. 2020. *Buku Panduan Indikator Kinerja Utama Perguruan Tinggi Negeri Edisi Tahun 2020*. Jakarta (ID).
- Fu L. 1994. *Neural Network in Computers Intelligence*. Singapura (SG): McGraw-Hill.
- Hustini Q. 2020. *Penerapan Chaid Untuk Klasifikasi Rasa Trauma Masyarakat Dki Jakarta Akibat Bencana Banjir* [skripsi]. Bogor (ID): Institut Pertanian Bogor.
- Hosmer DW, Lemeshow S. 2000. *Applied Logistic Regression*. Ed ke-2. New York [US] : John Wiley & Sons, Inc.
- Idham MA. 2020 Agus 17. *Daftar 15 Kampus Terbaik 2020 versi Kemendikbud: Skor IPB Tertinggi*. Tirto.id. Diunduh 2021 Jan 16. <https://tirto.id/daftar-15-kampus-terbaik-2020-versi-kemendikbud-skor-ipb-tertinggi-fYxF>.
- Institut Pertanian Bogor (IPB). 2019. *Jumlah Mahasiswa IPB University 2011-2019*. Ipb.ac.id. Diunduh 2021 Juli 8. <https://infografis.ipb.ac.id/index.php/jumlah-mahasiswa-ipb-university-2011-2019/>
- Kass GV. 1980. An Exploratory Technique for Investigating Large Quantities of Categorical Data. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, vol. 29, no. 2, 1980, pp. 119–127.
- Murphy E L, Comiskey C M. 2013. Using chi-Squared Automatic Interaction Detection (CHAID) modelling to identify groups of methadone treatment clients experiencing significantly poorer treatment outcomes. *Journal of Substance Abuse Treatment*. Volume 45, Issue 4.
- Oswald-Egg M E, Renold U. 2021. No experience, no employment: The effect of vocational education and training work experience on labour market outcomes after higher education. *Economics of Education Review*.
- Pertiwi RA. 2013. *Analisis CHAID Untuk Identifikasi Ketepatan Waktu Lulus Berdasarkan Karakteristik Mahasiswa* [skripsi]. Bogor (ID): Institut Pertanian Bogor.
- Putri YE. 2019. *Faktor-Faktor Yang Memengaruhi Status Pengangguran Di Provinsi Riau Dengan Metode Chaid* [skripsi]. Bogor (ID): Institut Pertanian Bogor.
- Kementerian Pendidikan dan Kebudayaan (Kemendikbud). 2021. *Tentang Kami Merdeka Belajar*. kampusmerdeka.kemdikbud.go.id. Diunduh 2021 Feb 22. <https://kampusmerdeka.kemdikbud.go.id/about/latar-belakang>.
- Kim H, & Loh W. 2001. Classification Trees with Unbiased Multiway Splits. *Journal of the American Statistical Association*, 96(454), 589-604.

- Kunto YS, Hasana SN. 2006. Analisis CHAID sebagai alat bantu statistika untuk segmentasi pasar (studi kasus pada koperasi syari'ah al-hidayah). *Jurnal Manajemen Pemasaran*. 1(2):88–98.
- Rianto H, Wahono RS. 2015. Resampling logistic regression untuk penanganan ketidakseimbangan class pada prediksi cacat software. *Journal of Software Engineering*. 1(1):46-53.
- Rokach, L. and Maimon, O., 2008. *Data Mining with Decision Tree*. USA: World Scientific.
- Roulin, N. and Bangerter, A., 2013. Students' use of extra-curricular activities for positional advantage in competitive job markets. *Journal of Education and Work*, 26(1), pp.21-47.
- Sepuh TJ. 2020 Aug 17. Direktorat Jenderal Pendidikan Tinggi Umumkan Klasterisasi Perguruan Tinggi Indonesia tahun 2020. dikti.kemdikbud.go.id. Diunduh 2021 Feb 1. <https://dikti.kemdikbud.go.id/kabar-dikti/kabar/direktorat-jenderal-pendidikan-tinggi-umumkan-klasterisasi-perguruan-tinggi-indonesia-tahun-2020/>.
- Spelmen VS, Porkodi R. 2018. A review on handling imbalanced data. Di dalam: 2018 International Conference on Current Trends towards Converging Technologies (ICCTCT); 2018 Mar 1-3; Coimbatore: IEEE. hlm 1-12. doi: 10.1109/ICCTCT.2018.8551020.
- Yap BW, Rani KA, Rahman HAA, Fong S, Khairudin Z, Abdullah NN. 2014. An application of *oversampling*, undersampling, bagging and boosting in handling imbalanced datasets. Di dalam: Herawan T, Deris MD, Abawajy J, editor. *Proceedings of the First International Conference on Advanced Data and Information Engineering (DaEng-2013)*; 2013 Des 13; Singapore, Singapore. Springer. hlm 13-22. doi:10.1007/978-981-4585-18-7_2.