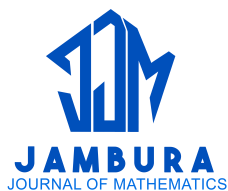


Penerapan Metode I-CHAID Menggunakan SMOTE pada Data Tidak Seimbang untuk Klasifikasi Durasi Studi Mahasiswa

Umar D. Akor, Muhammad Rezky Friesta Payu, dan La Ode Nashar



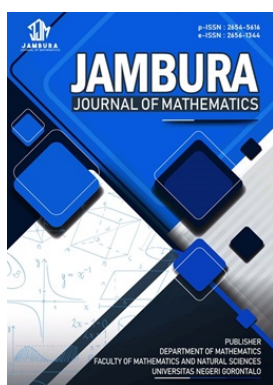
Volume 7, Issue 1, Pages 19–27, February 2025

Diterima 21 Oktober 2024, Direvisi 11 Desember 2024, Disetujui 20 Desember 2024, Diterbitkan 1 Februari 2025

To Cite this Article : U. D. Akor, M. R. F. Payu, dan L. O. Nashar, "Penerapan Metode I-CHAID Menggunakan SMOTE pada Data Tidak Seimbang untuk Klasifikasi Durasi Studi Mahasiswa", *Jambura J. Math*, vol. 7, no. 1, pp. 19–27, 2025, <https://doi.org/10.37905/jjom.v7i1.27978>

© 2025 by author(s)

JOURNAL INFO • JAMBURA JOURNAL OF MATHEMATICS

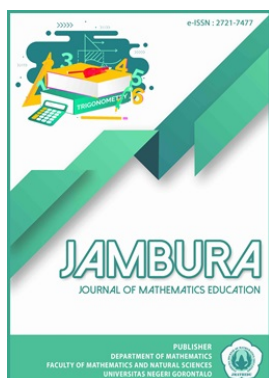


	Homepage	:	http://ejurnal.ung.ac.id/index.php/jjom/index
	Journal Abbreviation	:	Jambura J. Math.
	Frequency	:	Biannual (February and August)
	Publication Language	:	English (preferable), Indonesia
	DOI	:	https://doi.org/10.37905/jjom
	Online ISSN	:	2656-1344
	Editor-in-Chief	:	Hasan S. Panigoro
	Publisher	:	Department of Mathematics, Universitas Negeri Gorontalo
	Country	:	Indonesia
	OAI Address	:	http://ejurnal.ung.ac.id/index.php/jjom/oai
	Google Scholar ID	:	iWLjgaUAAAAJ
	Email	:	info.jjom@ung.ac.id

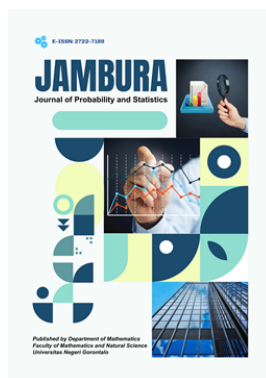
JAMBURA JOURNAL • FIND OUR OTHER JOURNALS



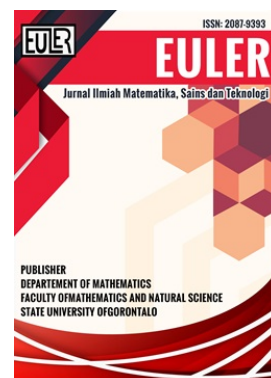
Jambura Journal of Biomathematics



Jambura Journal of Mathematics Education



Jambura Journal of Probability and Statistics



EULER : Jurnal Ilmiah Matematika, Sains, dan Teknologi



Penerapan Metode I-CHAID Menggunakan SMOTE pada Data Tidak Seimbang untuk Klasifikasi Durasi Studi Mahasiswa

Umar D. Akor^{1,*}, Muhammad Rezky Friesta Payu¹, dan La Ode Nashar¹

¹Program Studi Statistika, Universitas Negeri Gorontalo, Bone Bolango, Indonesia

ARTICLE HISTORY

Diterima 21 Oktober 2024
Direvisi 11 Desember 2024
Disetujui 20 Desember 2024
Diterbitkan 1 Februari 2025

KATA KUNCI

Durasi Studi
Klasifikasi
I-CHAID
SMOTE

KEYWORDS

Study Duration
Classification
Improved CHAID
SMOTE

ABSTRAK. Masalah keterlambatan kelulusan mahasiswa sering kali ditemui di berbagai perguruan tinggi, hal itu disebabkan oleh berbagai faktor baik faktor internal atau faktor eksternal. Faktor-faktor tersebut perlu untuk diklasifikasikan sehingga mudah untuk melakukan identifikasi terhadap faktor yang paling dominan. Metode klasifikasi yang digunakan pada penelitian ini yaitu Improved Chi-Square Automatic Interaction Detection (I-CHAID) dengan pendekatan Synthetic Minority Oversampling Technique (SMOTE). I-CHAID bertujuan untuk menentukan faktor yang paling dominan dalam membentuk pohon klasifikasi berdasarkan nilai chi-square dan tschuprow's, sedangkan SMOTE digunakan untuk mengatasi data yang jumlah amatannya tidak seimbang. Berdasarkan pohon klasifikasi I-CHAID dengan pendekatan SMOTE, diperoleh hasil signifikan yang mempengaruhi durasi penyelesaian studi yaitu IPK pada semester lima (67,20%) dan cara pembimbingan (87,50%). Adapun kinerja klasifikasi yang dihasilkan dari testing data 40% yaitu akurasi sebesar 40,60% atau dari 32 sampel terdapat 13 sampel yang tepat diklasifikasikan. Presentase sensitivitas yang dihasilkan yaitu 6,25% menunjukkan tingkat keberhasilan klasifikasi data dengan kategori lulus tepat waktu. Presentase spesifisitas yaitu 75% menunjukkan tingkat keberhasilan klasifikasi data dengan kategori lulus tidak tepat waktu. Presentase precision yaitu sebesar 20% menunjukkan hasil prediksi kategori lulus tepat waktu yang sebenarnya. Presentase F-measure yaitu 9,52% menunjukkan tingkat keseimbangan antara precision dan sensitivitas.

ABSTRACT. The issue of delayed graduation is often encountered in various universities, including in the Statistics Study Program at Universitas Negeri Gorontalo, for graduates between 2018 and 2023. Among them, 162 students (76.50%) experienced delayed graduation, and 5 students (2.35%) dropped out. This delay in graduation is caused by various factors, necessitating a classification method capable of identifying the most dominant factors. The classification method used in this research is Improved Chi-Square Automatic Interaction Detection (I-CHAID) with the Synthetic Minority Oversampling Technique (SMOTE) approach. SMOTE is employed to address imbalanced data. Based on the I-CHAID classification tree with the SMOTE approach, the significant factors influencing the duration of study completion are the GPA in the fifth semester (67.20%) and the mentoring method (87.50%). As for the classification performance from the 40% testing data, the accuracy achieved was 40.60%, meaning that out of 32 samples, 13 were correctly classified. The sensitivity value was 6.25%, indicating the success rate of classifying data for students who graduated on time. The specificity value was 75%, showing the success rate in classifying data for students who did not graduate on time. The precision value was 20%, reflecting the accuracy of predicting students who actually graduated on time, and the F-measure was 9.52%, indicating the balance between precision and sensitivity.



This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution-NonCommercial 4.0 International License. Editorial of JJoM: Department of Mathematics, Universitas Negeri Gorontalo, Jln. Prof. Dr. Ing. B. J. Habibie, Bone Bolango 96554, Indonesia.

1. Pendahuluan

Pendidikan merupakan kebutuhan utama bagi setiap individu, karena pendidikan memiliki kaitan yang erat dengan pengembangan kemampuan sumber daya manusia (SDM) yang diperoleh melalui proses pengajaran [1]. Meskipun pada faktanya, pendidikan dapat mengembangkan potensi seseorang akan tetapi sebagian orang tidak mampu menyelesaikan pendidikannya dengan tepat waktu. Pendidikan terdiri dari pendidikan informal yang diperoleh melalui lingkungan keluarga, pendidikan non formal diperoleh dari lingkungan sekitar, sedangkan pendidikan formal diperoleh dengan menempuh serangkaian tingkatan pendidikan

dari sekolah dasar hingga perguruan tinggi [2]. Perguruan tinggi memiliki tujuan untuk mencetak sarjana yang handal dan terampil di bidangnya [3]. Dengan kerjasama antar civitas akademika di program studi maka kompetensi lulusan dapat terpenuhi dan mahasiswa lulus tepat waktu akan terwujud [4]. Mahasiswa sarjana (S1) dikatakan lulus tepat waktu jika mampu menyelesaikan studinya sejak awal masuk perguruan tinggi hingga yudisium dengan waktu maksimal empat tahun [5]. Universitas Negeri Gorontalo terus berupaya untuk menekan angka lulusan tidak tepat waktu, sebagaimana yang tertuang dalam SK Rektor UNG tahun 2020 tentang peraturan akademik UNG. Penerapan aturan tersebut ditekan sampai pada tingkat fakultas, salah satunya fakultas MIPA UNG, khususnya Prodi Statistika.

*Penulis Korespondensi.

Berdasarkan data jumlah mahasiswa pada program studi Statistika yang diperoleh dari basis data akademik Universitas Negeri Gorontalo, ditemukan mahasiswa yang terlambat menyelesaikan studi dengan jumlah 162 orang (76,05%) dan mahasiswa *Drop Out* dengan jumlah 5 orang (2,35%), dari total mahasiswa yang telah menyelesaikan studi. Pada umumnya masalah keterlambatan kelulusan mahasiswa tersebut disebabkan oleh faktor ekonomi dan faktor suka menunda-nunda. Sulasteri dkk. [6] menyatakan bahwa penyebab keterlambatan penyelesaian tugas akhir mahasiswa dipengaruhi oleh faktor internal (kecerdasan, motivasi dan aktivitas) dan faktor eksternal (teman sebaya, dosen pembimbing, keluarga dan fasilitas). Faktor-faktor penyebab keterlambatan kelulusan tersebut perlu diklasifikasikan, sehingga faktor-faktor yang dikelompokkan dan membentuk kategori tersebut dapat dipahami karakteristiknya. Hasil klasifikasi tersebut dapat dijadikan sebagai acuan program studi dan solusi atas keterlambatan mahasiswa dalam menyelesaikan studi.

Dalam penerapan klasifikasi terdapat beberapa metode yang bisa digunakan diantaranya, *Naive Bayes* yaitu metode untuk mengklasifikasikan berdasarkan pada teorema bayes, dimana dapat menangani data dengan jumlah besar dan setiap fitur nya bersifat independen satu sama lain [7]. Metode lainnya yaitu *SVM(Support Vector Machine)* metode ini menggabungkan klasifikasi dan regresi dan bekerja dengan mencari *hyperplane* optimal yang memisahkan kelas-kelas data dalam ruang berdimensi tinggi [8]. Penelitian lainnya menggunakan metode K-Nearest Neighbor dan Random Forest untuk klasifikasi lama studi mahasiswa [9]. Sementara itu, metode *I-CHAID* merupakan metode pohon keputusan yang menggunakan uji *chi-square* dan uji *tshuprow's* untuk membagi data berdasarkan variabel yang paling signifikan [10]. Berdasarkan uraian diatas penulis menggunakan metode *I-CHAID* dalam melakukan klasifikasi faktor penyebab keterlambatan kelulusan mahasiswa.

Pengujian terhadap faktor-faktor penyebab keterlambatan kelulusan mahasiswa pada penelitian ini menggunakan metode klasifikasi *I-CHAID (Improved Chi-Squared Automatic Interaction Detection)*. *I-CHAID* adalah pengembangan dari metode *CHAID* yang mengkolaborasi nilai *Chi-square* dengan *Tschuprow's T* [11]. *Output* dari metode *I-CHAID* berupa grafis yang dihasilkan dari pohon klasifikasi, sehingga memudahkan dalam interpretasinya karena pola dari pemisahan dan penggabungan variabelnya dapat dilihat secara langsung [10]. Dalam pengaplikasian metode *I-CHAID* di temukan ketidakseimbangan data hasil observasi penelitian. Dimana, perbandingan antara jumlah mahasiswa lulus tepat waktu dan lulus tidak tepat waktu jauh berbeda. Sehingga, dibutuhkan suatu pendekatan yaitu dengan menggunakan metode *SMOTE (Synthetic Minority Oversampling Technique)*. Hal itu selaras dengan pernyataan Barro dkk. [12] yang mengatakan *SMOTE* merupakan metode yang digunakan untuk menangani permasalahan pada kategori data yang jumlah amatannya tidak seimbang dengan cara membangkitkan data sintetis pada kelas minor sehingga jumlah datanya menjadi setara dengan kelas mayor.

Penerapan metode *I-CHAID* dan penggunaan *SMOTE* telah dilakukan di beberapa penelitian, diantaranya penelitian mengenai metode *Improved Chi-Square Automatic Interaction Detection* oleh Bula dkk. [10] pada kasus literasi informasi dan penelitian Yuliasari dkk. [11] mengenai klasifikasi rumah tangga miskin di kecamatan kaubun tahun 2020. Selain itu terdapat penelitian

yang menggunakan pendekatan *SMOTE*, seperti penelitian yang dilakukan Erlin dkk. [13] mengenai dampak *SMOTE* terhadap kinerja *Random Forest Classifier* berdasarkan data tidak seimbang. Kemudian penelitian yang dilakukan oleh Santi dkk. [14] yaitu mengenai penerapan metode *SMOTE CHAID* dalam melakukan klasifikasi terhadap pasien tuberkulosis relapse. Berdasarkan penelitian terdahulu yang telah disebutkan diatas, maka peneliti melakukan kombinasi terhadap metode *I-CHAID* dan *SMOTE* dalam melakukan klasifikasi.

Penelitian ini dilakukan dengan mengkombinasikan metode *I-CHAID* dan *SMOTE*, pada penelitian ini variabel durasi penyelesaian studi mahasiswa merupakan variabel dependen, sedangkan untuk variabel independen terdapat sebanyak tujuh belas variabel. Adapun hasil penelitian dapat di jadikan sebagai acuan dalam mengatasi masalah keterlambatan kelulusan mahasiswa khususnya pada program studi Statistika, UNG. Tujuan penelitian yaitu untuk mengetahui variabel signifikan yang mempengaruhi durasi penyelesaian studi mahasiswa dan membentuk pohon faktor berdasarkan nilai variabel paling signifikan.

2. Metode

Metode yang digunakan dalam penelitian ini yaitu *I-CHAID* dengan pendekatan *SMOTE*. Data yang digunakan pada penelitian ini merupakan data kelulusan mahasiswa program studi Statistika, jurusan Matematika, Universitas Negeri Gorontalo angkatan 2018 dan 2019. Berdasarkan data tersebut, terdapat 81 sampel responden yang terbagi kedalam kategori lulus tepat waktu sebanyak 13 mahasiswa (16,04%) dan lulus tidak tepat waktu sebanyak 68 mahasiswa (83,95%).

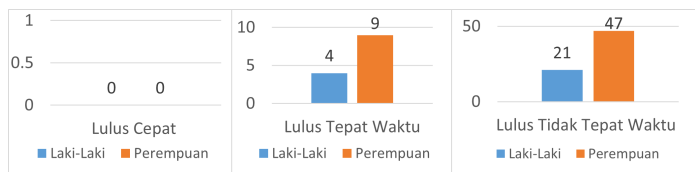
Tahapan analisis data dalam penelitian ini adalah sebagai berikut:

1. *Input* data penelitian
2. Membuat diagram frekuensi antara mahasiswa lulus tepat waktu dan mahasiswa lulus tidak tepat waktu dan melakukan eksplorasi data dengan menggunakan diagram batang dan diagram lingkaran dan membandingkan data untuk melihat karakteristik responden.
3. *Splitting data* menjadi data *training* dan data *testing* dengan pembagian rasio data 60% untuk data *training* dan 40% untuk data *testing* [11].
4. Menerapkan *SMOTE* untuk menyeimbangkan data dengan membentuk data sintetis menggunakan data *training* 60%.
5. Melakukan uji *Chi-Square* berdasarkan data hasil *SMOTE* untuk mengetahui faktor-faktor yang mempengaruhi durasi penyelesaian studi.
6. Melakukan uji *Tschuprow's* terhadap hasil uji *Chi-Square*.
7. Melakukan koreksi bonferroni terhadap hasil uji *Tschuprow's*
8. Melakukan pemisahan yaitu *split node* terhadap *training data* berdasarkan hasil uji *Tschuprow's* yang telah terkoreksi bonferroni untuk membentuk pohon klasifikasi.
9. Melakukan penghentian pembentukan pohon klasifikasi setelah mencapai batas akhir pembentukan *node*.
10. Melakukan pengujian kinerja klasifikasi berdasarkan data *testing* 40% menggunakan *confusion matrix* (akurasi, sensitivitas dan spesifisitas) dan *F-measure*.
11. Interpretasi model.

3. Hasil dan Pembahasan

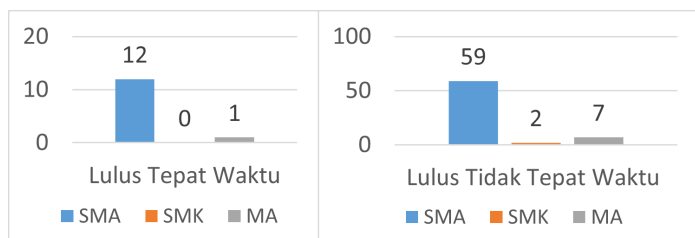
3.1. Eksplorasi Data

Eksplorasi data menggunakan diagram batang dan diagram lingkaran untuk membandingkan data dan menampilkan karakteristik responden terkait keterlambatan penyelesaian studi mahasiswa Proghram Studi Statistika, Universitas Negeri Gorontalo, khususnya angkatan 2018 dan 2019. Hasil eksplorasi data disajikan pada Gambar 1-18.



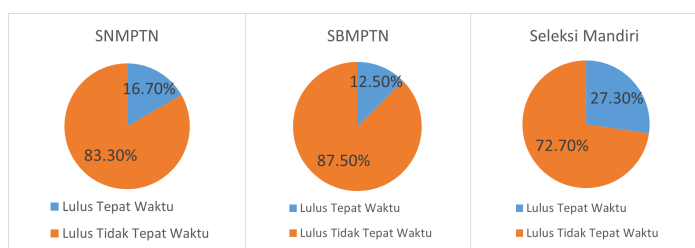
Gambar 1. Deskripsi durasi penyelesaian studi berdasarkan jenis kelamin

Berdasarkan Gambar 1, dapat ditunjukkan bahwa mahasiswa mendominasi durasi penyelesaian studi untuk jenis kelamin perempuan pada kategori lulus tepat waktu dan kategori lulus tidak tepat waktu. Selain itu, terlihat bahwa tidak terdapat mahasiswa yang lulus cepat waktu.



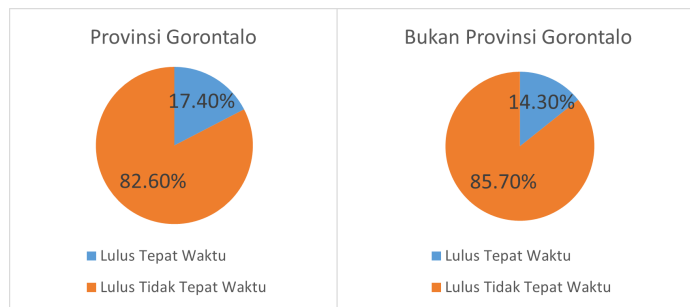
Gambar 2. Deskripsi durasi penyelesaian studi berdasarkan asal sekolah

Berdasarkan Gambar 2, dapat ditunjukkan bahwa mahasiswa yang berasal dari sekolah SMA paling banyak lulus tepat waktu namun juga termasuk kategori yang lulus tidak tepat waktu. Selain itu, terlihat bahwa mahasiswa yang berasal dari SMK merupakan yang paling sedikit pada setiap kategori.



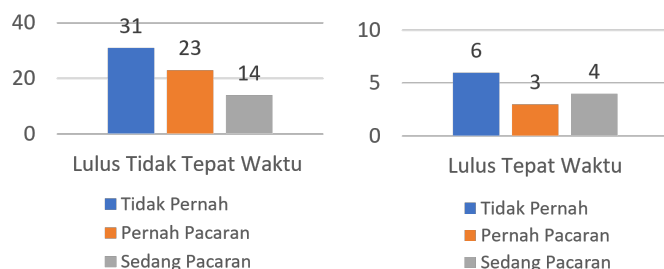
Gambar 3. Deskripsi durasi penyelesaian studi berdasarkan jalur masuk perguruan tinggi

Berdasarkan Gambar 3, terlihat bahwa mahasiswa yang masuk perguruan tinggi berdasarkan jalur SNMPTN, SBMPTN dan seleksi mandiri merupakan yang 37 mendominasi pada kategori lulus tidak tepat waktu. Selain itu, terlihat bahwa mahasiswa yang masuk perguruan tinggi berdasarkan jalur seleksi mandiri memiliki presentase terbesar pada kategori lulus tepat waktu dan mahasiswa yang masuk perguruan tinggi berdasarkan jalur SBMPTN memiliki presentase terkecil pada kategori lulus tepat waktu.



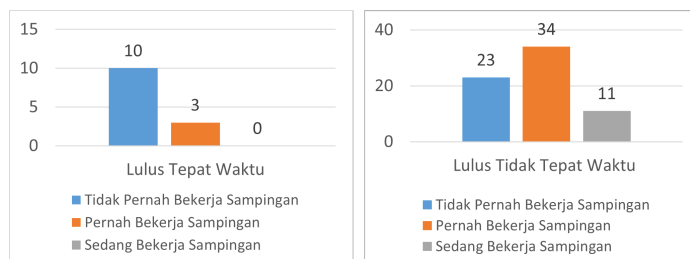
Gambar 4. Deskripsi durasi penyelesaian studi berdasarkan asal daerah

Berdasarkan Gambar 4, terlihat bahwa mahasiswa yang berasal dari provinsi gorontalo atau bukan provinsi gorontalo mendominasi kategori lulus tidak tepat waktu. Selain itu, terlihat bahwa mahasiswa yang berasal dari provinsi gorontalo memiliki presentase terbesar pada kategori lulus tepat waktu.



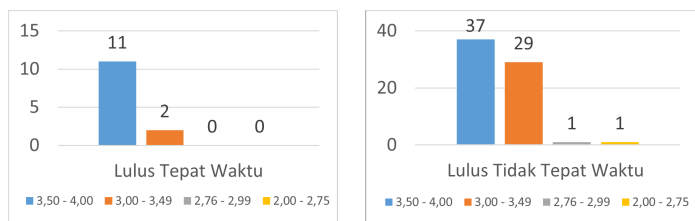
Gambar 5. Deskripsi durasi penyelesaian studi berdasarkan status pacaran

Berdasarkan Gambar 5, dapat ditunjukkan bahwa mahasiswa yang tidak pernah pacaran mendominasi durasi penyelesaian studi pada kategori lulus tepat waktu dan kategori lulus tidak tepat waktu. Selain itu, terlihat bahwa mahasiswa yang pernah pacaran merupakan yang paling sedikit pada kategori lulus tepat waktu dan mahasiswa yang sedang pacaran merupakan yang paling sedikit pada kategori lulus tidak tepat waktu.



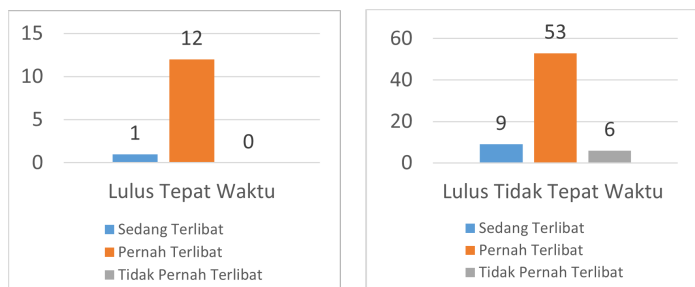
Gambar 6. Deskripsi durasi penyelesaian studi berdasarkan pekerjaan sampingan

Berdasarkan Gambar 6, dapat ditunjukkan bahwa mahasiswa mendominasi durasi penyelesaian studi untuk mahasiswa yang tidak pernah bekerja sampingan pada kategori lulus tepat waktu sedangkan yang mendominasi kategori lulus tidak tepat waktu adalah mahasiswa yang pernah bekerja sampingan. Selain itu, terlihat bahwa mahasiswa yang sedang bekerja sampingan merupakan yang paling sedikit pada setiap kategori.



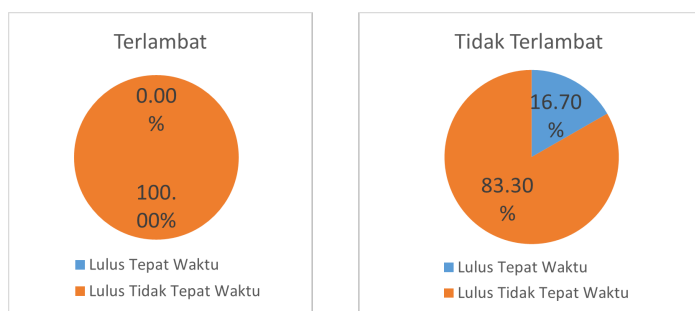
Gambar 7. Deskripsi durasi penyelesaian studi berdasarkan IPK pada semester lima (5)

Berdasarkan Gambar 7, dapat ditunjukkan bahwa mahasiswa mendominasi durasi penyelesaian studi untuk nilai IPK pada semester lima pada kategori lulus tepat waktu dan kategori lulus tidak tepat waktu. Selain itu, terlihat bahwa mahasiswa dengan rentang nilai antara 2,00–2,75 dan mahasiswa dengan rentang nilai 2,76– 2,99 merupakan yang paling sedikit pada setiap kategori.



Gambar 8. Deskripsi durasi penyelesaian studi berdasarkan keterlibatan dalam organisasi

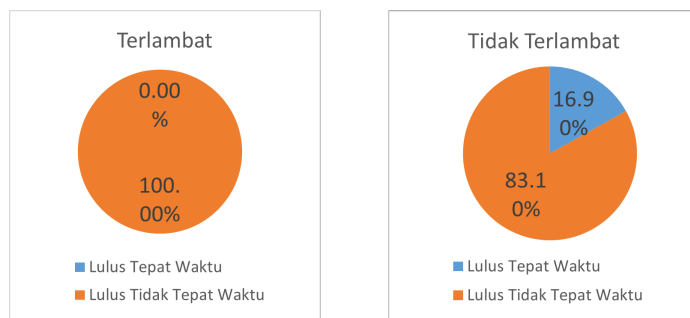
Berdasarkan Gambar 8, dapat ditunjukkan bahwa mahasiswa mendominasi durasi penyelesaian studi untuk mahasiswa yang pernah terlibat dalam organisasi pada kategori lulus tepat waktu dan kategori lulus tidak tepat waktu. Selain itu, terlihat bahwa mahasiswa yang tidak pernah terlibat dalam organisasi merupakan yang paling sedikit pada setiap kategori.



Gambar 9. Deskripsi durasi penyelesaian studi berdasarkan magang

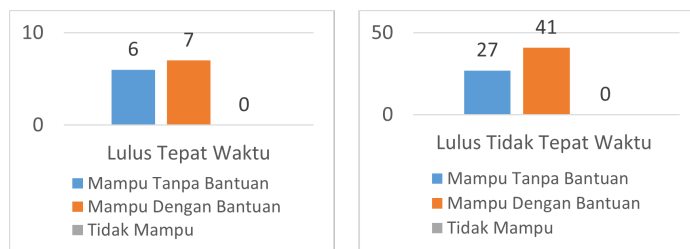
Berdasarkan Gambar 9, terlihat bahwa tidak terdapat mahasiswa dengan kategori lulus tidak tepat waktu yang terlambat magang. Selain itu, terlihat bahwa mahasiswa yang tidak terlambat magang merupakan yang paling sedikit pada kategori lulus tepat waktu.

Berdasarkan Gambar 10, terlihat bahwa tidak terdapat mahasiswa dengan kategori lulus tidak tepat waktu yang terlambat KKN. Selain itu, terlihat bahwa mahasiswa yang tidak terlambat



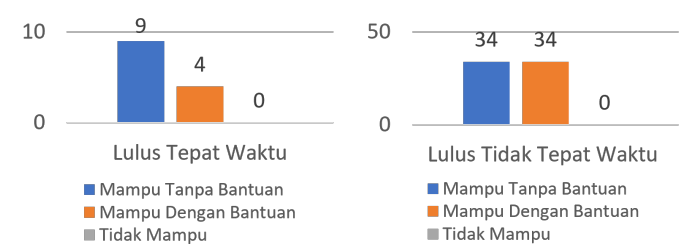
Gambar 10. Deskripsi durasi penyelesaian studi berdasarkan KKN

KKN merupakan yang paling sedikit pada kategori lulus tepat waktu.



Gambar 11. Deskripsi durasi penyelesaian studi berdasarkan kemampuan mengoperasikan software statistika

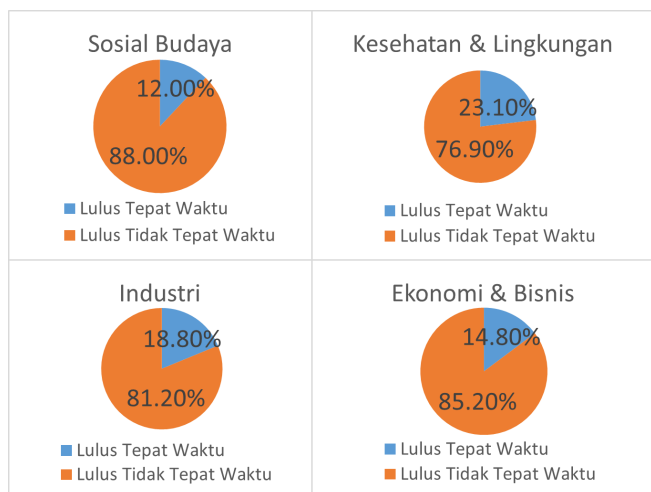
Berdasarkan Gambar 11, dapat ditunjukkan bahwa mahasiswa mendominasi durasi penyelesaian studi untuk mahasiswa yang mampu mengoperasikan software untuk analisis data namun tetap dibantu orang lain pada kategori lulus tepat waktu dan kategori lulus tidak tepat waktu. Selain itu, terlihat bahwa mahasiswa yang mampu mengoperasikan software untuk analisis data tanpa dibantu orang lain merupakan yang paling sedikit pada setiap kategori.



Gambar 12. Deskripsi Durasi Penyelesaian Studi Berdasarkan Kemampuan mengoperasikan LATEX

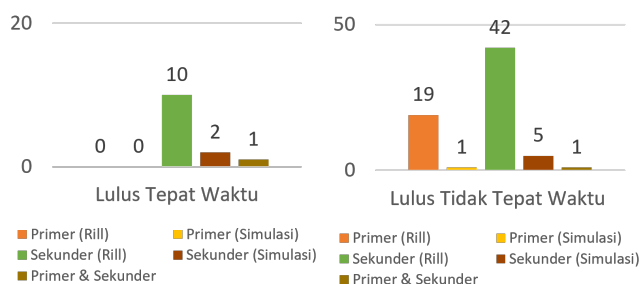
Berdasarkan Gambar 12, dapat ditunjukkan bahwa mahasiswa mendominasi durasi penyelesaian studi untuk mahasiswa yang mampu mengoperasikan LATEX untuk penulisan naskah skripsi tanpa dibantu orang lain pada kategori lulus tepat waktu. Selain itu, terlihat bahwa mahasiswa yang mampu mengoperasikan LATEX untuk penulisan naskah skripsi tanpa dibantu orang lain seimbang dengan mahasiswa yang mampu mengoperasikan LATEX untuk penulisan naskah skripsi namun tetap dibantu orang lain.

Berdasarkan Gambar 13, terlihat bahwa mahasiswa dengan kajian terapan studi kasus sosial budaya, ekonomi dan bisnis, in-



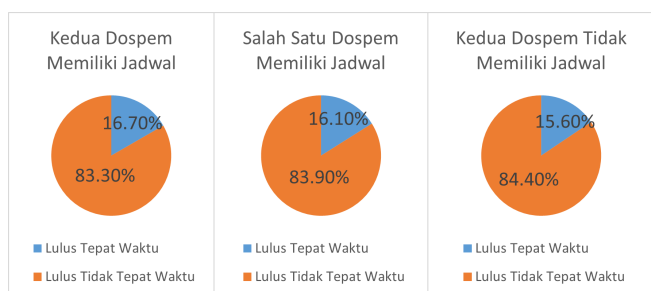
Gambar 13. Deskripsi durasi penyelesaian studi berdasarkan kajian terapan studi kasus

dustri dan kesehatan dan lingkungan mendominasi lulus tidak tepat waktu pada setiap kategori. Selain itu, terlihat bahwa mahasiswa dengan kajian terapan studi kasus kesehatan dan lingkungan memiliki presentase terbesar pada kategori lulus tepat waktu dan mahasiswa dengan kajian terapan studi kasus sosial budaya memiliki presentase terkecil pada kategori lulus tepat waktu.



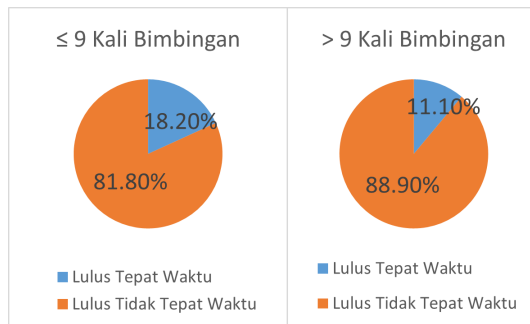
Gambar 14. Deskripsi durasi penyelesaian studi berdasarkan sumber data penelitian

Berdasarkan Gambar 14, dapat ditunjukkan bahwa mahasiswa mendominasi durasi penyelesaian studi untuk mahasiswa dengan sumber data penelitian sekunder (rill) pada kategori lulus tepat waktu dan kategori lulus tidak tepat waktu. Selain itu, terlihat bahwa tidak ada mahasiswa yang menggunakan sumber data primer (rill) dan sumber data primer (simulasi) pada lulus tepat waktu.



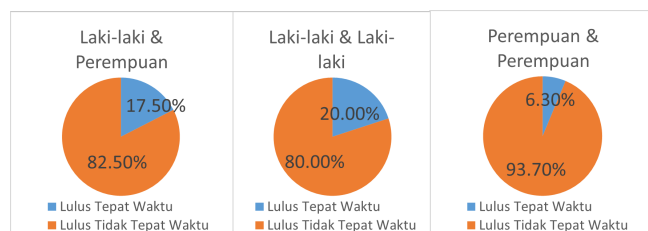
Gambar 15. Deskripsi durasi penyelesaian studi berdasarkan cara pembimbingan

Berdasarkan Gambar 15, terlihat bahwa mahasiswa dengan cara pembimbingan terjadwal memiliki presentase terbesar pada kategori lulus tepat waktu sedangkan cara pembimbingan tidak terjadwal memiliki presentase terkecil pada kategori lulus tepat waktu. Selain itu, terlihat bahwa yang mendominasi merupakan kategori lulus tidak tepat waktu.



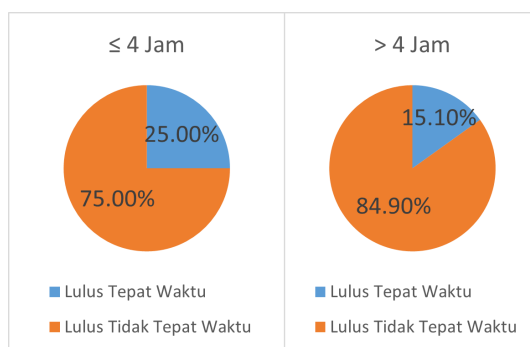
Gambar 16. Deskripsi durasi penyelesaian studi berdasarkan rata-rata durasi bimbingan

Berdasarkan Gambar 16, terlihat bahwa mahasiswa yang durasi bimbingan kurang dari sembilan kali atau lebih dari sembilan kali mendominasi kategori lulus tidak tepat waktu. Selain itu, terlihat bahwa mahasiswa dengan durasi bimbingan kurang dari sembilan kali memiliki presentase terbesar pada kategori lulus tepat waktu.



Gambar 17. Deskripsi durasi penyelesaian studi berdasarkan jenis kelamin dosen pembimbing utama

Berdasarkan Gambar 17, terlihat bahwa mahasiswa dengan jenis kelamin dosen pembimbing utamanya laki-laki dan laki-laki memiliki presentase terbesar pada kategori lulus tepat waktu sedangkan yang jenis kelamin dosen pembimbing utamanya perempuan dan perempuan memiliki presentase terkecil pada kategori lulus tepat waktu. Selain itu, terlihat bahwa yang mendominasi merupakan kategori lulus tidak tepat waktu.



Gambar 18. Deskripsi durasi penyelesaian studi berdasarkan durasi penggunaan gawai (gadget)

Tabel 1. Hasil uji Chi-Square

Pasangan Variabel	χ_{hit}^2	χ_{tabel}^2	p-value	Keterangan
Jenis Kelamin dan Durasi penyelesaian studi	0,9425	3,841	0,3316	Tidak Signifikan
Asal sekolah dan Durasi penyelesaian studi	0	3,841	1	Tidak Signifikan
Jalur Masuk perguruan tinggi dan Durasi penyelesaian studi	2,49	5,991	0,2879	Tidak Signifikan
Asal daerah dan Durasi penyelesaian studi	5,256	3,841	0,0219	Signifikan
Status Pacaran Selama Studi dan Durasi penyelesaian studi	5,9024	3,841	0,0151	Signifikan
Pekerjaan Sampingan dan Durasi penyelesaian studi	11,215	3,841	0,0008	Signifikan
IPK semester 5 dan Durasi penyelesaian studi	28,23	3,841	$1,47 \times 10^{-8}$	Signifikan
Keterlibatan dalam Organisasi dan Durasi penyelesaian studi	10,11	3,841	0,001	Signifikan
Terlambat Magang dan Durasi penyelesaian studi	3,114	3,841	0,1200	Tidak Signifikan
Terlambat KKN dan Durasi penyelesaian studi	4,206	3,841	0,0578	Tidak Signifikan
Kemampuan mengoperasikan software statistikan untuk analisis data dan Durasi penyelesaian studi	0,053	3,841	0,8179	Tidak Signifikan
Kemampuan mengoperasikan LATEX untuk penulisan naskah skripsi dan Durasi penyelesaian studi	9,848	3,841	0,0017	Signifikan
Kajian terapan studi kasus skripsi dan Durasi penyelesaian studi	2,609	5,991	0,2713	Tidak Signifikan
Sumber data penelitian dan Durasi penyelesaian studi	0	3,841	1	Tidak Signifikan
Cara pembimbingan dan Durasi penyelesaian studi	13,309	5,991	0,0013	Signifikan
Rata-rata durasi bimbingan skripsi dan Durasi penyelesaian studi	0,3059	3,841	0,5802	Tidak Signifikan
Jenis kelamin dosen pembimbing dan Durasi penyelesaian studi	1,246	3,841	0,2643	Tidak Signifikan

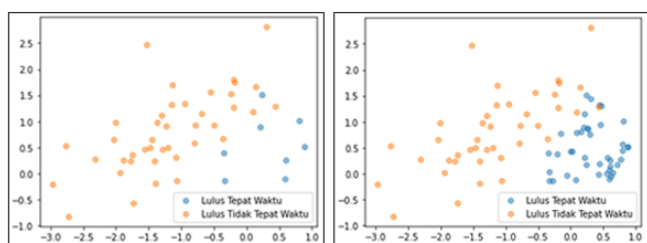
Berdasarkan Gambar 18, terlihat bahwa mahasiswa yang durasi penggunaan gawai kurang dari empat jam per hari atau lebih dari empat jam per hari mendominasi kategori lulus tidak tepat waktu. Selain itu, terlihat bahwa mahasiswa dengan durasi penggunaan gawai kurang dari empat jam per hari memiliki presentase terbesar pada kategori lulus tepat waktu.

3.2. Splitting Data

Splitting data bertujuan untuk memisahkan data penelitian menjadi data latih (training data) dan data uji (testing data). Training data digunakan untuk menyeimbangkan data dengan teknik SMOTE dan juga pengujian melalui analisis I-CHAID, sedangkan testing data digunakan untuk menguji kinerja klasifikasi yang telah dibentuk menggunakan data latih. Adapun pembagian data yaitu menggunakan rasio 60% untuk data latih dan 40% untuk data uji [11].

3.3. Membangkitkan Data Sintetis Menggunakan SMOTE

Teknik SMOTE digunakan untuk menangani ketidakseimbangan kelas dalam set data dengan cara membentuk data sintetis [15]. Adapun hasil SMOTE ditampilkan pada Gambar 19.



(a). Data Sebelum Smote (b). Data Sesudah Smote

Gambar 19. Hasil SMOTE

Berdasarkan Gambar 19, dapat tunjukan bahwa data pada gambar (a) merupakan data sebelum dilakukan SMOTE dan datanya belum seimbang. Sedangkan data pada gambar (b) merupakan data setelah dilakukan SMOTE dan datanya telah seimbang.

Setelah data tersebut seimbang maka selanjutnya yang akan dilakukan yaitu uji independensi dengan menggunakan uji chi square.

3.4. Uji Chi-Square Menggunakan Data Hasil SMOTE

Uji chi-square digunakan untuk menentukan apakah terdapat hubungan independen antara dua variabel. Uji independensi yang digunakan yaitu uji koreksi yates, uji fisher exact test dan uji pearson chi-square. Pasangan variabel dianggap signifikan jika p-value lebih kecil dari tingkat signifikan alpha (α) 0,05 atau 5%, atau jika nilai χ_{hit}^2 melebihi nilai χ_{tab}^2 , Pasangan variabel yang paling signifikan ditentukan berdasarkan nilai p-value terkecil atau nilai χ_{hit}^2 terbesar [16]. Adapun persamaan uji chi-square yang digunakan pada penelitian ini merujuk pada [11] dan dinyatakan sebagai berikut:

1. Kontingensi berukuran 2×2

$$\chi_{hit}^2 = \frac{N(O_{11} \times O_{22} - O_{12} \times O_{21})^2}{n_{.1} \times n_{.2} \times n_{.1} \times n_{.2}} \tag{1}$$

2. Kontingensi berukuran $B \times K$

$$\chi_{hit}^2 = \sum_{ij=1}^N \left[\frac{(O_{ij} - E_{ij})^2}{E_{ij}} \right] \tag{2}$$

dengan

$$E_{ij} = \left(\frac{n_{i.} \times n_{.j}}{N} \right) \tag{3}$$

Keterangan:

- O_{ij} : Jumlah pengamatan baris ke- i dan kolom ke- j ,
- E_{ij} : Nilai harapan pada baris ke- i dan kolom ke- j ,
- $n_{i.}$: Nilai total pada kolom ke- j ,
- $n_{.j}$: Nilai total pada baris ke- i ,
- N : Nilai total seluruh pengamatan,
- i : Jumlah baris,
- j : Jumlah kolom.

Selanjutnya, hasil uji chi-square pada training data 60% ditunjukkan pada Tabel 1.

Setelah ditemukan hasil uji *chi square* yang signifikan, langkah selanjutnya yaitu melakukan uji *tshuprow's* berdasarkan hasil uji *chi-square*.

3.5. Uji Tshuprow's

Uji *Tshuprow's* dilakukan untuk melihat tingkat hubungan pada masing-masing pasangan variabel yang dinyatakan signifikan pada uji *Chi-Square*. *Tshuprow's* dalam Balasubramaniam dan Gunasundari [17] secara umum menyatakan persamaan Uji *Tshuprow's* sebagai berikut:

$$T = \frac{\chi^2}{W \sqrt{(V - 1)(U - 1)}} \tag{4}$$

dengan

V : Kategori variabel independen,

U : Kategori variabel dependen,

W : Jumlah sampel.

Selanjutnya, untuk mengetahui kuat lemahnya tingkat keamatan hubungan antara variabel perlu dilakukan pengkategorian dengan kategorisasi nilai korelasi merujuk pada [18] dan dinyatakan pada **Tabel 2**.

Tabel 2. Kategorisasi nilai korelasi

Besarnya Nilai Hubungan (Koefisien Korelasi)	Keterangan
$0, 50 \leq KK < 1, 00$	Kuat
$0, 20 \leq KK < 0, 50$	Sedang
$0, 00 \leq KK < 0, 20$	Lemah

Berdasarkan **Tabel 1**, hasil uji *chi-square* menunjukkan bahwa X_4 (Variabel Asal Daerah), X_5 (Status Pacaran Selama Studi), X_6 (Pekerjaan Sampingan), X_7 (IPK pada semester lima), X_8 (Keterlibatan Dalam Organisasi), X_{12} (Kemampuan Mengoperasikan LATEX) dan X_{15} (Cara Pembimbingan) memiliki hubungan yang signifikan terhadap Durasi Penyelesaian Studi Mahasiswa. Variabel X_6 (IPK pada semester lima) memiliki nilai tertinggi yaitu $\chi_{hit}^2 = 28, 23$ dan $p\text{-value} = 1,47 \times 10^{-8}$. Hasil uji *chi-square* signifikan dilanjutkan dengan langkah uji *tshuprow's* yang dapat dilihat pada **Tabel 3**.

Tabel 3. Hasil uji *Tshuprow's*

Pasangan Variabel	Tshuprow's	Keterangan
Y dan X_4	0,064	Lemah
Y dan X_5	0,021	Lemah
Y dan X_6	0,228	Sedang
Y dan X_7	0,344	Sedang
Y dan X_8	0,123	Lemah
Y dan X_{12}	0,12	Lemah
Y dan X_{15}	0,114	Lemah

Tabel 3 menunjukkan bahwa pasangan variabel independen (Asal Daerah (X_4), Status Pacaran Selama Studi (X_5), Keterlibatan Dalam Organisasi (X_8), Kemampuan Mengoperasikan LATEX (X_{12}) dan Cara Pembimbingan (X_{15})) memiliki hubungan lemah terhadap variabel Durasi Penyelesaian Studi. Sedangkan untuk variabel independen (Pekerjaan Sampingan (X_6) dan IPK pada semester 5 (X_7)) memiliki hubungan sedang terhadap variabel Durasi Penyelesaian Studi dengan nilai *tshuprow's* terbesar adalah variabel IPK pada semester 5 yaitu 0,344.

3.6. Koreksi Bonferroni

Koreksi bonferroni dilakukan terhadap variabel signifikan yang telah dilakukan penggabungan kategori pada langkah uji *chi-square* [19]. Hasil perhitungan alfa yang terkoreksi *Bonferro-ni* dibandingkan dengan nilai *p-value*, jika nilai *p-value* lebih kecil dari nilai *Bonferro-ni* maka nilai *p-value* dianggap signifikan. Pada penelitian ini diperoleh nilai alfa terkoreksi *Bonferro-ni* yaitu $\alpha_{Bonferro-ni} = \frac{0,05}{5} = 0,01$. Karena hasil alfa terkoreksi *Bonferro-ni* lebih kecil dari nilai *p-value* (0,05), maka disimpulkan bahwa seluruh variabel signifikan tetap digunakan dalam pembentukan *split node*.

3.7. Pemisahan Node Pohon Klasifikasi

Tahap pemisahan dilakukan untuk tujuan pembentuk pohon klasifikasi berdasarkan *split node*. Dalam menentukan *split node* didasarkan pada besaran nilai koefisien korelasi uji *tshuprow's*, pasangan variabel yang signifikan dengan nilai *tshuprow's* tertinggi yaitu IPK pada semester 5 (0,344). Hal itu menunjukkan bahwa IPK pada semester 5 akan dijadikan pemisah dari variabel durasi penyelesaian studi. Langkah selanjutnya menentukan *internal node* yang akan menjadi pemisah dari variabel IPK pada semester 5 dengan perhitungan yang sama seperti dalam pemilihan *split node*.

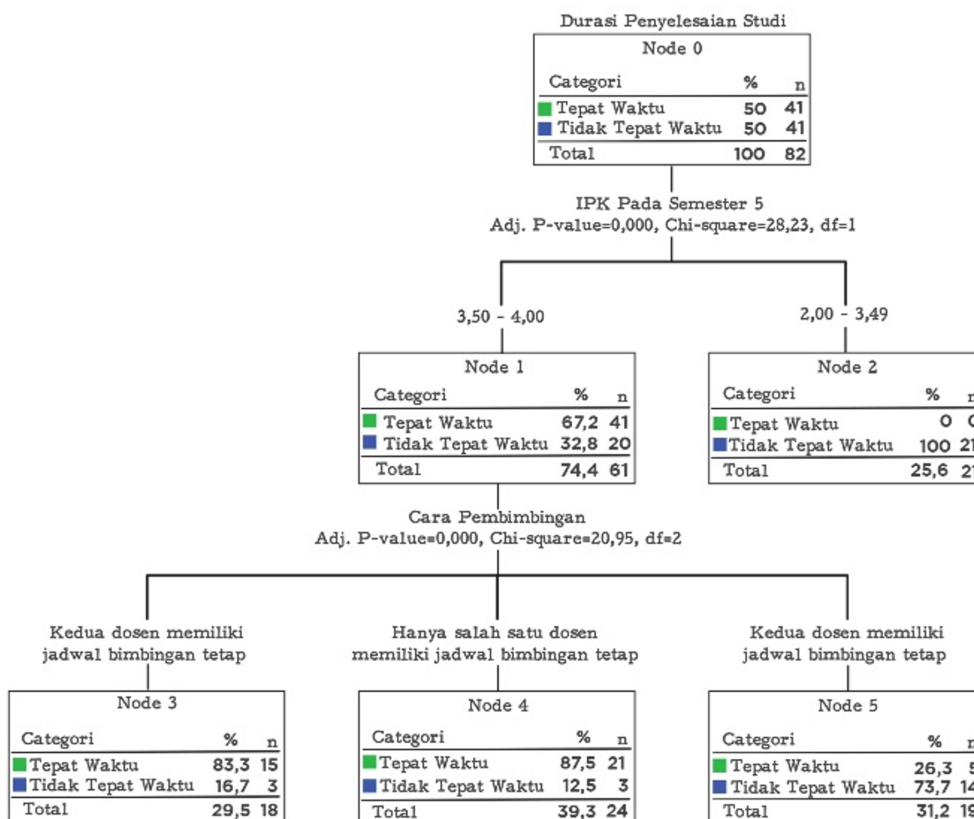
Pemisahan untuk variabel IPK pada semester 5 yaitu variabel cara pembimbingan, dengan hasil uji *tshuprow's* sebesar 0,3434, yang ditentukan berdasarkan nilai cabang dari variabel IPK pada semester 5. Karena tidak ditemukan lagi variabel yang signifikan setelah dilakukan pengujian selanjutnya pada variabel Asal Daerah, Status Pacaran Selama Studi, Keterlibatan Dalam Organisasi dan Kemampuan Mengoperasikan LATEX, maka tahap *splitting node* dihentikan.

3.8. Penghentian Pohon Klasifikasi

Penghentian adalah proses untuk menentukan *leaf node* atau batas akhir dalam pembentukan pohon klasifikasi. Pembentukan pohon klasifikasi dihentikan ketika *internal node* tidak dapat dibentuk lagi atau telah memenuhi aturan pertumbuhan pohon. Hasil klasifikasi I-CHAID dengan pendekatan SMOTE menggunakan *training data* 60% ditampilkan pada **Gambar 20**.

Gambar 20 menunjukkan jumlah sampel dengan data latih 60% sebanyak 82 responden. *Root node* pohon klasifikasi yaitu variabel Y (Durasi Penyelesaian Studi) dengan dua kategori yaitu lulus tepat waktu sebesar 41 (50%) responden dan lulus tidak tepat waktu sebesar 41 (50%) responden. Pada gambar juga dapat dilihat terdapat dua variabel independen yang signifikan yakni IPK pada semester 5 dan Cara Pembimbingan, pembentukan pohon klasifikasi tersebut menghasilkan 4 segmen yaitu sebagai berikut:

1. Dari 82 sampel terdapat 21 sampel yang memiliki nilai IPK pada semester 5 dengan rentang nilai antara 2,00 – 3,49 yang dibagi berdasarkan dua kategori durasi penyelesaian studi yaitu lulus tepat waktu sebanyak (0%) atau 0 sampel dan lulus tidak tepat waktu sebanyak (100%) atau 21 sampel.
2. Dari 82 Sampel terdapat 61 sampel yang memiliki nilai IPK pada semester 5 dengan rentang nilai antara 3,50 – 4,00 yang dibagi berdasarkan dua kategori durasi penyelesaian studi yaitu lulus tepat waktu sebanyak (67,20%) atau 41 sampel dan lulus tidak tepat waktu sebanyak (32,80%) 51



Gambar 20. Pohon klasifikasi I-CHAID menggunakan SMOTE

atau 20 sampel. Tetapi dari 61 sampel tersebut terdapat 19 sampel cara pembimbingan kedua dosen tidak memiliki jadwal bimbingan tetap yang dibagi berdasarkan dua kategori durasi penyelesaian studi yaitu lulus tepat waktu sebanyak (26,30%) atau 5 sampel dan lulus tidak tepat waktu sebanyak (73,70%) atau 14 sampel.

3. Dari 82 Sampel terdapat 61 sampel yang memiliki nilai IPK pada semester 5 dengan rentang nilai antara 3,50 – 4,00 yang dibagi berdasarkan dua kategori durasi penyelesaian studi yaitu lulus tepat waktu sebanyak (67,20%) atau 41 sampel dan lulus tidak tepat waktu sebanyak (32,80%) atau 20 sampel. Tetapi dari 61 sampel tersebut terdapat 24 sampel cara pembimbingan hanya salah satu dosen memiliki jadwal bimbingan tetap yang dibagi berdasarkan dua kategori durasi penyelesaian studi yaitu lulus tepat waktu sebanyak (87,50%) atau 21 sampel dan lulus tidak tepat waktu sebanyak (12,50%) atau 3 sampel.
4. Dari 82 Sampel terdapat 61 sampel yang memiliki nilai IPK pada semester 5 dengan rentang nilai antara 3,50 – 4,00 yang dibagi berdasarkan dua kategori durasi penyelesaian studi yaitu lulus tepat waktu sebanyak (67,20%) atau 41 sampel dan lulus tidak tepat waktu sebanyak (32,80%) atau 20 sampel. Tetapi dari 61 sampel tersebut terdapat 18 sampel cara pembimbingan hanya salah satu dosen memiliki jadwal bimbingan tetap yang dibagi berdasarkan dua kategori durasi penyelesaian studi yaitu lulus tepat waktu sebanyak (83,30%) atau 15 sampel dan lulus tidak tepat waktu sebanyak (16,70%) atau 3 sampel.

3.9. Kinerja Klasifikasi

Kinerja klasifikasi merupakan proses evaluasi terhadap hasil klasifikasi yang telah dibentuk sebelumnya. Pengujian klasifikasi dilakukan menggunakan *testing data* untuk mengukur ketepatan hasil klasifikasi berdasarkan *training data*. Statistik uji yang digunakan yaitu *confusion matrix* guna mengevaluasi akurasi, sensitivitas dan spesifisitas, serta *F-measure* yang bertujuan untuk melihat keseimbangan antara presisi dan sensitivitas [20]. Hasil uji kinerja klasifikasi dengan *testing data* 40% dapat di lihat pada Tabel 4.

Tabel 4. Hasil kinerja klasifikasi

Uji	Hasil
Akurasi	40%
Sensitivitas	6,25%
Spesifisitas	75%
<i>Precision</i>	20%
<i>F – Measure</i>	9,25%

Berdasarkan Tabel 4, dapat diketahui hasil dari kinerja klasifikasi pada testing data 40% memiliki nilai akurasi 40,60% yang memiliki arti dari 32 sampel terdapat 13 sampel tepat untuk diklasifikasikan. Adapun nilai sensitivitas yang menunjukkan tingkat keberhasilan kinerja klasifikasi *Improved CHAID* menggunakan pendekatan *SMOTE* untuk mengklasifikasi data dengan kategori lulus tepat waktu sebesar 6,25%, nilai spesifitas yang dihitung menunjukkan keberhasilan kinerja klasifikasi *Improved CHAID* menggunakan pendekatan *SMOTE* untuk mengklasifikasi data dengan kategori lulus tidak tepat waktu sebesar 75%, nilai *pre-*

cision yang dihitung menunjukkan keberhasilan kinerja klasifikasi *Improved CHAID* menggunakan pendekatan *SMOTE* sebesar 20% untuk memprediksi data dengan kategori lulus tepat waktu sebenarnya dari total kelas lulus tepat waktu, sedangkan nilai *F-measure* menunjukkan keseimbangan antara nilai *Precision* dan nilai *Sensitivitas* sebesar 9,52%.

4. Kesimpulan

Faktor-faktor signifikan yang mempengaruhi durasi penyelesaian studi mahasiswa program studi Statistika, Universitas Negeri Gorontalo berdasarkan hasil analisis *I-CHAID* dengan pendekatan *SMOTE* yaitu IPK pada semester lima (67,20%) dan cara pembimbingan (87,50%). Hasil klasifikasi analisis *Improved CHAID* dengan pendekatan *SMOTE* menggunakan data latih 60% menghasilkan empat segmen yaitu IPK pada semester lima pada rentang nilai 2,00 – 3,49, kedua dosen pembimbing memiliki jadwal bimbingan tetap, salah satu dosen pembimbing memiliki jadwal bimbingan tetap dan kedua dosen pembimbing tidak memiliki jadwal bimbingan tetap. Selain itu, juga menghasilkan tiga variabel independen yaitu variabel IPK pada semester lima dan variabel cara pembimbingan. Hasil kinerja klasifikasi dengan data uji 40% menghasilkan akurasi sebesar 40,60% atau dari 32 sampel terdapat 13 sampel tepat diklasifikasikan. Nilai *sensitivitas* yaitu 6,25% menunjukkan tingkat keberhasilan klasifikasi data dengan kategori lulus tepat waktu. Nilai *spesifisitas* 75% menunjukkan tingkat keberhasilan klasifikasi data dengan kategori lulus tidak tepat waktu. Nilai *precision* 20% menunjukkan hasil prediksi kategori lulus tepat waktu yang sebenarnya. Adapun nilai *F-measure* 9,52% menunjukkan tingkat keseimbangan antara *precision* dan *sensitivitas*.

Kontribusi Penulis. Umar D. Akor: Perangkat lunak, sumber daya, kurasi data, penulisan—penyusunan draf awal, visualisasi. Muhammad Rezky Friesta Payu: Metodologi, Konseptualisasi, analisis formal, supervisi. La Ode Nashar: Penulisan—tinjauan dan pengeditan, validasi, supervisi. Semua penulis telah membaca dan menyetujui versi manuskrip yang diterbitkan.

Ucapan Terima Kasih. Para penulis menyampaikan terima kasih kepada editor dan reviewer atas pembacaan yang cermat, kritik yang mendalam, dan rekomendasi yang praktis untuk meningkatkan kualitas tulisan ini.

Pembiayaan. Penelitian ini tidak menerima pembiayaan eksternal.

Konflik Kepentingan. Para penulis menyatakan tidak ada konflik kepentingan yang terkait dengan artikel ini.

Referensi

- [1] F. Nofratama, H. Hasrul, H. Muchtar, and S. F. Dewi, "Kendala Keterlambatan Penyelesaian Studi Mahasiswa PPKn FIS Universitas Negeri Padang," *J. Educ. Cult. Polit.*, vol. 2, no. 2, pp. 185–191, Nov. 2022, doi: [10.24036/jecc.v2i2.106](https://doi.org/10.24036/jecc.v2i2.106).
- [2] A. Adawiyah, S. Sulfasyah, and J. Arifin, "Implikasi Pendidikan Nonformal Pada Remaja," *Equilib. J. Pendidik. Sociol.*, vol. 4, no. 2, pp. 1–8, Feb. 2017, doi: [10.26618/equilibrium.v4i2.506](https://doi.org/10.26618/equilibrium.v4i2.506).
- [3] M. Annur, J. A. Dahlan, and F. Agustina, "Penerapan Metode Multivariate Adaptive Regression Spline (MARS) untuk Menentukan Faktor yang Mempengaruhi Masa Studi Mahasiswa FPMIPA UPI," *J. EurekaMatika*, vol. 3, no. 1, pp. 135–155, 2015, [Online]. Available: <https://dspace.uui.ac.id/123456789/27720>.
- [4] I. G. A. M. Srinadi and D. P. E. Nilakusumawati, "Analisis Waktu Kelulusan Mahasiswa FMIPA Universitas Udayana dan Faktor-faktor yang Mempengaruhinya," *E-Jurnal Mat.*, vol. 9, no. 3, p. 205, Sep. 2020, doi: [10.24843/MTK.2020.v09.i03.p300](https://doi.org/10.24843/MTK.2020.v09.i03.p300).
- [5] N. D. Larasati and W. S. Jatiningrum, "Analisis Faktor pada Keterlambatan Studi Mahasiswa Teknik Industri Universitas Ahmad Dahlan," *Manaj. Pendidik.*, vol. 16, no. 2, pp. 83–96, Dec. 2021, doi: [10.23917/jmp.v16i2.12134](https://doi.org/10.23917/jmp.v16i2.12134).
- [6] S. Sulasteri, F. Nur, and A. Kusumayanti, "Faktor-faktor penyebab keterlambatan mahasiswa UIN Alauddin Makassar menyelesaikan skripsi," *J. Idaarah*, vol. III, no. 1, pp. 96–113, 2019, doi: [10.24252/idaarah.v3i1.9389](https://doi.org/10.24252/idaarah.v3i1.9389).
- [7] S. Telaumbanua, S. A. Abimannyu, and G. Y. Abdullah, "Literatur Review: Pendekatan Naïve Bayes Untuk Klasifikasi Penyakit Tanaman Kentang," *BIIKMA Bul. Ilm. Ilmu Komput. dan Multim.*, vol. 2, no. 3, pp. 420–428, 2024, [Online]. Available: <https://jurnalmahasiswa.com/index.php/biikma/article/view/1603>.
- [8] E. Novianto, S. Suhirman, and D. Prasetyo, "Perbandingan metode klasifikasi random forest dan support vector machine dalam memprediksi capaian studi mahasiswa," *JUPI (Jurnal Ilm. Penelit. dan Pembelajaran Inform.)*, vol. 9, no. 4, pp. 1821–1833, 2024, doi: [10.29100/jupi.v9i4.5423](https://doi.org/10.29100/jupi.v9i4.5423).
- [9] I. K. Hasan, R. Resmawan, and J. Ibrahim, "Perbandingan K-Nearest Neighbor dan Random Forest dengan Seleksi Fitur Information Gain untuk Klasifikasi Lama Studi Mahasiswa," *Indones. J. Appl. Stat.*, vol. 5, no. 1, pp. 58–66, 2022, doi: [10.13057/ijas.v5i1.58056](https://doi.org/10.13057/ijas.v5i1.58056).
- [10] Z. Bula, R. Resmawan, L. O. Nashar, and S. K. Nasib, "Implementasi Improved Chi-Square Automatic Interaction Detection Pada Klasifikasi Faktor-faktor yang Mempengaruhi Literasi Informasi Generasi Muda," *J. Stat. dan Apl.*, vol. 6, no. 2, pp. 214–222, Dec. 2022, doi: [10.21009/JSA.06207](https://doi.org/10.21009/JSA.06207).
- [11] P. D. Yuliasari, R. Goejantoro, and F. D. T. Amijaya, "Klasifikasi Rumah Tangga Miskin Di Kecamatan Kaubun Tahun 2020 Dengan Menggunakan Metode Improved Chi-Square Automatic Interaction Detection," *EKSPONENSIAL*, vol. 12, no. 1, p. 83, Jun. 2021, doi: [10.30872/eksponensial.v12i1.765](https://doi.org/10.30872/eksponensial.v12i1.765).
- [12] R. A. Barro, I. D. Sulvianti, and F. M. Afendi, "Penerapan Synthetic Minority Oversampling Technique (SMOTE) Terhadap Data Tidak Seimbang Pada Pembuatan Model Komposisi Jamu," *Xplore J. Stat.*, vol. 1, no. 1, pp. 1–6, 2013, doi: [10.29244/xplore.v1i1.12424](https://doi.org/10.29244/xplore.v1i1.12424).
- [13] E. Erlin, Y. Desnelita, N. Nasution, L. Suryati, and F. Zoromi, "Dampak SMOTE terhadap Kinerja Random Forest Classifier berdasarkan Data Tidak seimbang," *MATRIK J. Manajemen, Tek. Inform. dan Rekayasa Komput.*, vol. 21, no. 3, pp. 677–690, Jul. 2022, doi: [10.30812/matrik.v21i3.1726](https://doi.org/10.30812/matrik.v21i3.1726).
- [14] V. M. Santi, L. Nafisah, and Q. Meidianingsih, "Penerapan Metode SMOTE CHAID dalam Klasifikasi Tuberkulosis Relapse," *J. Stat. dan Apl.*, vol. 6, no. 1, pp. 26–36, Jun. 2022, doi: [10.21009/JSA.06103](https://doi.org/10.21009/JSA.06103).
- [15] I. S. Hidayati and I. M. Arcana, "Penerapan CHAID Dengan Pendekatan SMOTE Pada Kematian Balita di Kawasan Timur Indonesia Tahun 2017," *Semin. Nas. Off. Stat.*, vol. 2019, no. 1, pp. 357–367, May 2020, doi: [10.34123/semnasoffstat.v2019i1.97](https://doi.org/10.34123/semnasoffstat.v2019i1.97).
- [16] I. C. Negara and A. Prabowo, "Penggunaan Uji Chi-Square untuk Mengetahui Pengaruh Tingkat Pendidikan dan Umur terhadap Pengetahuan Penasun Mengenai HIV-AIDS di Provinsi DKI Jakarta," *Pros. Semin. Nas. Mat. dan Ter.* 2018, vol. 1, no. 1, pp. 1–8, 2018.
- [17] C. P. Balasubramaniam and D. R. Gunasundari, "Supply Chain Enhancement Using Improved Chaid Algorithm for Classifying the Customer Groups," *Int. J. Comput. Appl. Technol. Res.*, vol. 6, no. 8, pp. 378–383, Aug. 2017, doi: [10.7753/IJCATR0608.1006](https://doi.org/10.7753/IJCATR0608.1006).
- [18] Y. Salsabila, A. Fatah, and J. Jaenudin, "Hubungan antara Literasi Numerasi terhadap Kemampuan Berpikir Kritis dan Kreatif Siswa SMP di Kecamatan Curug," *Equal. J. Ilm. Pendidik. Mat.*, vol. 6, no. 1, pp. 42–54, Jun. 2023, doi: [10.46918/equals.v6i1.1789](https://doi.org/10.46918/equals.v6i1.1789).
- [19] C. Damayanti, D. Kusnandar, and Yudhi, "Perbandingan Hasil Pembentukan Pohon Klasifikasi Metode CHAID dan Improved CHAID," *Bul. Ilm. Mat. Stat. dan Ter.*, vol. 07, no. 4, pp. 10–27, 2018, doi: [10.26418/bbimst.v7i4.28655](https://doi.org/10.26418/bbimst.v7i4.28655).
- [20] G. Zeng, "On the confusion matrix in credit scoring and its analytical properties," *Commun. Stat. - Theory Methods*, vol. 49, no. 9, pp. 2080–2093, May 2020, doi: [10.1080/03610926.2019.1568485](https://doi.org/10.1080/03610926.2019.1568485).