

PENINGKATAN ALGORITMA DECISION TREE DALAM MENGKLASIFIKASI TINGKAT KELULUSAN MAHASISWA UNIVERSITAS JAMBI DENGAN SELEKSI FITUR *CHI-SQUARE*

Reza Wahyu Hardian¹, Jasmir², Sharipuddin³

^{1,2,3} Fakultas Ilmu Komputer, Program Studi Magister Sistem Informasi, Universitas Dinamia Bangsa, Jambi, Indonesia

Email: ¹reza74204@gmail.com, ²jasmir@unama.ac.id, ³*sharipuddin@unama.ac.id

Email Penulis Korespondensi: reza74204@gmail.com

Submitted :
17 April 2025

Revision :
24 Juni 2025

Accepted:
25 September 2025

Published:
30 September 2025

Abstrak—Peningkatan jumlah mahasiswa Universitas Jambi setiap tahunnya tidak diiringi dengan tingkat kelulusan yang sebanding, sehingga diperlukan upaya untuk menganalisis data mahasiswa guna memberikan solusi strategis. Penelitian ini bertujuan untuk melakukan analisis data mining pada data mahasiswa untuk membantu pembimbing akademik memprediksi status kelulusan mahasiswa dan memberikan peringatan dini agar mahasiswa dapat menyelesaikan studi tepat waktu, sehingga angka keterlambatan kelulusan dapat diminimalkan. Data yang digunakan dalam penelitian ini adalah data alumni Fakultas Sains dan Teknologi Universitas Jambi tahun 2019 hingga 2024, yang telah melalui proses data cleaning. Metode klasifikasi yang digunakan adalah algoritma Decision Tree dengan penerapan seleksi fitur Chi-Square untuk meningkatkan akurasi model. Hasil penelitian menunjukkan bahwa penerapan seleksi fitur pada algoritma Decision Tree menghasilkan tingkat akurasi sebesar 80,00%, lebih tinggi dibandingkan model tanpa seleksi fitur yang hanya mencapai akurasi sebesar 78,57%. Selain itu, nilai presisi meningkat dari 86,82% menjadi 84,41%, recall meningkat dari 86,72% menjadi 92,24%, dan F1-score meningkat dari 86,77% menjadi 88,13%. Temuan ini menunjukkan bahwa seleksi fitur berkontribusi signifikan dalam meningkatkan kinerja model klasifikasi untuk memprediksi kelulusan mahasiswa Universitas Jambi, terutama dalam meningkatkan recall yang mencerminkan kemampuan model dalam mengidentifikasi mahasiswa yang lulus tepat waktu secara lebih akurat.

Kata Kunci: Data Mining; Machine Learning; Klasifikasi; Decision Tree; Chi-Square

Abstract— The increasing number of students at Universitas Jambi each year is not accompanied by a proportional graduation rate, making it necessary to analyze student data to provide strategic solutions. This study aims to apply data mining techniques to student data to assist academic advisors in predicting students' graduation status and providing early warnings to help students complete their studies on time, thereby minimizing graduation delays. The data used in this study consists of alumni records from the Faculty of Science and Technology at Universitas Jambi from 2019 to 2024, which have undergone a data cleaning process. The classification method used is the Decision Tree algorithm with Chi-Square feature selection to enhance model accuracy. The results indicate that applying feature selection to the Decision Tree algorithm improves the accuracy to 80.00%, compared to 78.57% without feature selection. Additionally, the precision increased from 86.82% to 84.41%, recall improved from 86.72% to 92.24%, and F1-score rose from 86.77% to 88.13%. These findings suggest that feature selection significantly contributes to enhancing the classification model's performance in predicting student graduation at Universitas Jambi, particularly by improving recall, which reflects the model's ability to more accurately identify students who graduate on time.

Keywords: Data Mining; Machine Learning; Classification; Decision Tree; Chi-Square

1. PENDAHULUAN

Pendidikan tinggi, termasuk di Universitas Jambi (UNJA), memegang peranan strategis dalam pembangunan sumber daya manusia yang berkualitas. Keberhasilan mahasiswa dalam menyelesaikan studi tepat waktu merupakan salah satu indikator utama yang mencerminkan kualitas penyelenggaraan pendidikan di institusi tersebut. Berbagai faktor, seperti kemampuan akademik, latar belakang sosial-ekonomi, kehadiran, serta tingkat keterlibatan mahasiswa dalam kegiatan akademik maupun non-akademik, dapat mempengaruhi capaian kelulusan tepat waktu. Pemahaman yang komprehensif terhadap faktor-faktor tersebut memungkinkan institusi pendidikan, termasuk UNJA, untuk merumuskan strategi yang lebih efektif dalam meningkatkan tingkat kelulusan mahasiswa. Tingkat kelulusan tepat waktu juga sangat krusial bagi perguruan tinggi, karena menjadi salah satu tolok ukur dalam penilaian mutu dan peringkat institusi. Tingkat kelulusan merupakan aspek atau ukuran tingkat keberhasilan suatu perguruan tinggi dalam melaksanakan proses belajar mengajarnya [1]. Namun menurut Saputra dkk tidak semua mahasiswa dapat menyelesaikan studinya tepat waktu [2]. Oleh sebab itu, diperlukan suatu teknik untuk

memprediksi waktu kelulusan mahasiswa untuk memberikan rekomendasi kepada Universitas atau Fakultas atau Program Studi dalam mengurangi mahasiswa yang lulus tidak tepat waktu.

Lama studi merupakan atribut penting bagi suatu universitas baik itu fakultas maupun bagi program studi. Memprediksi lama studi mahasiswa dapat dilakukan dari beberapa faktor dan kriteria yang digunakan untuk mengukur kelulusan mahasiswa [3]. Evaluasi pengukuran dilakukan guna untuk menghasilkan sarjana yang memiliki daya saing yang tinggi dan berkualitas dengan memanfaatkan basis data akademik sebagai pendukung keputusan oleh manajemen perguruan tinggi dalam meningkatkan standart kualitas. Kemajuan teknologi di bidang data mining, kecerdasan buatan, dan pembelajaran mesin membuka peluang baru dalam analisis data pendidikan. Teknologi ini memungkinkan institusi memanfaatkan tumpukan data akademik untuk mengidentifikasi pola, tren, dan faktor-faktor yang memengaruhi lama studi mahasiswa. Dengan memanfaatkan teknik data mining, institusi dapat menghasilkan informasi berharga guna menunjang kegiatan manajemen, operasional harian, dan pengambilan keputusan strategis.

Pemanfaatan teknik data mining pada perguruan tinggi dengan memanfaatkan tumpukan data yang digunakan untuk mendapatkan informasi guna menunjang kegiatan manajemen dan kegiatan operational harian serta bermanfaat dalam mengambil keputusan. Salah satu teknik dalam data mining yaitu klasifikasi data menggunakan algoritma *Decision Tree*. Algoritma ini memiliki kemampuan untuk memproses data yang kompleks dan menghasilkan model yang dapat diinterpretasikan dengan mudah. *Decision Tree* merupakan model yang dapat digunakan dalam mengklasifikasi data yang dimana model ini menunjukkan informasi berharga mengenai faktor – faktor yang mempengaruhi kelulusan [4]. Hal ini memungkinkan penentuan pola dan tren yang relevan dengan tingkat kelulusan yang akan menghasilkan tingkat akurasi yang baik. Dalam meningkatkan tingkat akurasi dari kinerja model dapat dilakukan dengan melakukan seleksi fitur sebelum melakukan pemodelan, misalnya dengan menggunakan algoritma *chi-square* [5]. *Chi-Square* adalah salah satu algoritma yang dapat digunakan dalam meningkatkan tingkat akurasi dari kinerja permodelan dalam klasifikasi data. Uji *Chi-Square* sering diterapkan untuk mengevaluasi signifikansi suatu fitur dengan menghitung nilai statistik *Chi-Square* guna menentukan tingkat keterkaitannya dengan kelas.

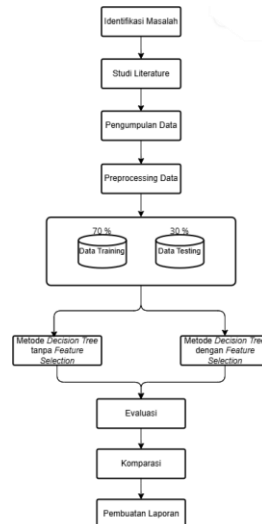
Pada penelitian yang dilakukan oleh Moerdyanto dan Nuryana pada tahun 2021 berjudul Prediksi Kelulusan Tepat Waktu Menggunakan Pendekatan Pohon Keputusan Algoritma *Decision Tree* yang dipublikasikan pada *Journal of Informatics and Computer Science*. Pada penelitian ini dilakukan klasifikasi data kelulusan mahasiswa prodi Sistem Informasi, jurusan Teknik Informatika, Universitas Negeri Surabaya menggunakan algoritma *Decision Tree* yang bertujuan untuk melakukan data mining pada riwayat akademik mahasiswa dengan harapan dapat mengantisipasi dan menangani dini kemungkinan mahasiswa yang tidak lulus atau lulus tepat waktu. Dari penelitian tersebut diperoleh tingkat akurasi sebesar 75,95% [6]. Penelitian yang dilakukan oleh Qisthiano dkk pada tahun 2023, dilakukan penelitian penerapan algoritma *Decision Tree* dalam klasifikasi data prediksi kelulusan mahasiswa dengan sampel yaitu data alumni dari beberapa peruruguan tinggi di Kota Palembang dengan jumlah sebesar 1739 dan dilakukan klasifikasi menggunakan tools *RapidMiner* [3]. Pada penelitian ini dilakukan klaisifikasi data menggunakan algoritma *Decision Tree* memperoleh tingkat akurasi sebesar 87,93%. Penelitian yang dilakukan oleh A.Saputra dkk pada tahun 2022 didalam penelitian Optimasi *Chi-Square* Dan Perbaikan Teknik *Prunning* Untuk Peningkatan Akurasi Algoritma C4.5 dalam model kasus prediksi keterlambatan biaya kuliah dilakukan klasifikasi data dengan tujuan untuk menentukan fitur, implementasi algoritma C4.5 dan mengevaluasi kinerja algoritma C4.5 [7]. Pada penelitian ini juga dilakukan penerapan *feature selection* menggunakan metode *Chi-Square* pada data yang berjumlah 12.408 data. Penelitian ini menghasilkan tingkat akurasi algoritma C4.5 tanpa *feature selection* sebesar 61,40% sedangkan menggunakan *feature selection* 65,53%. Penelitian yang dilakukan oleh Darmaputra dan Nugroho pada tahun 2019 didalam penelitian yang berjudul *Student Graduation Prediction Using Algoritma K-Means With Fitur Selection Chi-Square* dilakukan penelitian menggunakan algoritma K-Means dengan *fitur selection Chi-Square* dalam memprediksi masa studi mahasiswa dengan hasil penelitian didapatkan tingkat akurasi sebesar 86,57 % dengan pengaruh *fitur selection Chi-Square* terhadap algoritma K-Means sebesar 8,85% dalam meningkatkan akurasi dari kinerja [8].

Berdasarkan dari permasalahan di atas serta beberapa penelitian sejenis yang dimana algoritma *Decision Tree* dan atribut *Chi-Square* memiliki tingkat akurasi yang baik dalam melakukan klasifikasi data, maka penulis akan melakukan proses klasifikasi data menggunakan algoritma *Decision Tree* dan *feature selection* dengan metode *Chi-Square* dalam mengklasifikasi data tingkat kelulusan mahasiswa yang telah menempuh kuliah sarjana.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Pada penelitian klasifikasi data ini dalam mencapai tujuan penelitian dilakukan dengan algoritma *decision tree* dengan fitur seleksi *chi-square* yang dilakukan dalam beberapa tahapan, berikut tahapan yang dilakukan dalam penelitian dapat dilihat pada Gambar 1.



Gambar 1. Tahapan Penelitian

Berdasarkan kerangka kerja penelitian yang telah digambarkan, maka dapat diuraikan pembahasan masing-masing tahap dalam penelitian adalah sebagai berikut :

1. **Identifikasi Masalah**
 Pada tahap ini, penulis mengidentifikasi masalah yang terkait dengan permasalahan klasifikasi data mengenai masa studi mahasiswa di Fakultas Sains dan Teknologi, Universitas Jambi. Identifikasi ini bertujuan untuk menyusun rencana kerja serta menentukan jenis data yang akan diperlukan dalam penelitian ini.
2. **Studi Literatur**
 Pada tahap ini, penulis memperluas wawasan untuk menemukan topik yang layak dijadikan penelitian dengan mempelajari dan memahami berbagai teori dan konsep. Penulis melakukan kajian melalui buku, jurnal, makalah, serta berbagai sumber, termasuk dari Perpustakaan Universitas Dinamika Bangsa Jambi.
3. **Data Alumni**
 Pada tahap ini penulis melakukan pengumpulan data alumni dan informasi yang dibutuhkan dalam penelitian. Data yang akan digunakan merupakan data alumni Fakultas Sains dan Teknologi Universitas Jambi dari tahun 2020 hingga 2024.
4. **Preprocessing Data**
 Sebelum memulai proses penggalian informasi dalam klasifikasi data, dilakukan seleksi terhadap data operasional yang akan digunakan. Data yang dipilih adalah data alumni Fakultas Sains dan Teknologi Universitas Jambi. Sebelum pelaksanaan data mining, diperlukan proses pembersihan data yang menjadi fokus dalam *Knowledge Discovery in Database (KDD)*. Proses pembersihan ini meliputi penghapusan data duplikat, pemeriksaan data yang tidak konsisten, serta perbaikan kesalahan pada data, seperti kesalahan penulisan.
5. **Split Data**
 Data yang digunakan setelah melakukan proses persiapan data dengan perbandingan 70% untuk data training dalam membentuk model *Decision Tree* dan 30% untuk data *testing* yang digunakan untuk mengevaluasi hasil kinerja dari model *Decision Tree*. Menurut Joseph (2022) proses pembagian data 70:30 banyak digunakan dalam proses pembagian data dalam sebuah penelitian dan tidak ada penjelasan yang jelas mengenai rasio apa yang terbaik untuk kumpulan data tertentu [9].
6. **Klasifikasi dengan *Decision Tree* tanpa Seleksi Fitur *Chi-Square***
 Pada tahapan ini, dilakukan proses pengklasifikasian data dengan menggunakan algoritma decision tree tanpa menggunakan sekesi fitur dalam prose pemodelan. Kemudian proses klasifikasi dengan *Decision Tree* merupakan tahapan akhir yang dilakukan yaitu untuk mendapatkan hasil klasifikasi.
7. **Klasifikasi dengan *Decision Tree* dan Seleksi Fitur *Chi-Square***
 Pada tahapan ini, dilakukan seleksi fitur dengan *Chi-Square* terhadap data yang sudah terseleksi yang artinya data sudah siap untuk data mining dan dilakukan proses pengklasifikasian hasil dari seleksi fitur dengan *Chi-Square*. Kemudian proses klasifikasi dengan *Decision Tree* merupakan tahapan akhir yang dilakukan yaitu untuk mendapatkan hasil klasifikasi.
8. **Komparasi**
 Pada tahapan ini dilakukan proses perbandingan tingkat akurasi klasifikasi data menggunakan algoritma *Decision Tree* tanpa seleksi fitur dengan klasifikasi data menggunakan algoritma *Decision tree* degan seleksi fitur *Chi-Square*.

9. Evaluation

Pola informasi yang dihasilkan dari proses klasifikasi data harus disajikan dalam format yang mudah dipahami oleh pihak terkait. Tahap ini merupakan bagian dari proses *Knowledge Discovery in Database* (KDD) yang disebut interpretasi. Pada tahap ini, dilakukan juga pemeriksaan untuk memastikan apakah pola atau informasi yang ditemukan sesuai atau bertentangan dengan fakta dengan mengetahui tingkat akurasi model dalam mengklasifikasi data.

10. Pembuatan Laporan

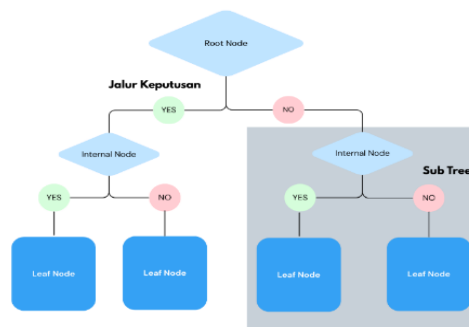
Pada tahap ini kesimpulan diambil berdasarkan hasil penelitian yang dibuat dalam bentuk laporan penelitian.

2.2 Decision Tree

Pohon keputusan (*Decision Tree*) merupakan salah satu metode yang banyak digunakan dalam klasifikasi dan prediksi. Metode ini sangat populer karena kemampuannya dalam mengkategorikan data dan membuat perkiraan berdasarkan aturan-aturan tertentu. Berikut ini akan dijelaskan beberapa istilah yang berkaitan dengan metode ini untuk memperjelas konsep yang digunakan dalam proses pengklasifikasian maupun prediksi. Menurut Sriani dkk., "*Decision Tree are a classification technique in data mining that is useful for creating tree structures to produce answers to problems that have been entered*" [10].

Decision Tree juga disebut sebagai diagram alir yang berbentuk seperti struktur pohon yang mana setiap internal node menyatakan pengujian terhadap suatu atribut, setiap cabang menyatakan output dari pengujian tersebut dan node daun (*leaf node*) menyatakan distribusi kelas. Node yang paling atas disebut sebagai node akar (*root node*). *Decision Tree* digunakan untuk mengklasifikasikan suatu sampel data yang belum diketahui kelasnya ke dalam kelas-kelas yang sudah ada. Jalur pengujian data adalah pertama semua data harus melalui *root node* dan terakhir adalah melalui *leaf node* yang akan menyimpulkan prediksi kelas bagi data tersebut. Atribut data harus berupa data kategorik, bila kontinu maka atribut harus didiskretisasi terlebih dahulu. *Decision Tree* merupakan salah satu algoritma dalam pembelajaran mesin yang sering digunakan untuk melakukan tugas-tugas klasifikasi maupun regresi. Struktur *Decision Tree* dimulai dari node, cabang dan keputusan [11].

Decision Tree



Gambar 2. *Decision Tree*

a. Node

Node terbagi menjadi tiga tingkatan, yaitu *Root Node*, *Internal Node*, dan *Leaf Node*. *Root Node* merupakan *node* teratas pada struktur pohon yang mewakili keseluruhan dataset.

b. Cabang

Cabang berfungsi sebagai penghubung antar *node* dan menunjukkan hasil evaluasi dari suatu atribut.

c. Keputusan

Keputusan ditentukan dengan mengikuti jalur dari *Root Node* menuju *Leaf Node* pada pohon keputusan.

Langkah – langkah dalam proses klasifikasi pada algoritma *Decision Tree* sebagai berikut.

a. Pilih Atribut Terbaik

Dalam menentukan atribut terbaik untuk *node*, digunakan beberapa kriteria seperti *Gini Impurity*, *Entropy*, dan *Information Gain* untuk mengukur kualitas atribut tersebut. Rumus menghitung nilai *Entropy* dan *Information Gain* sebagai berikut.

$$Entropy(S) = \sum_{i=1}^n - p_i * \log_2 p_i \quad (1)$$

$$Gain = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy(S) \quad (2)$$

Keterangan:

S = himpunan kasus

n = jumlah partisi atribut A

Pi = proporsi Si terhadap S

$|S_i|$ = jumlah kasus pada partisi ke i

$|S|$ = jumlah kasus dalam S

A = atribut

- b. Pembagian Dataset
Dataset dipecah menjadi beberapa subset berdasarkan nilai atribut yang telah dipilih sebelumnya.
- c. Pembangunan Pohon
Proses ini dilakukan secara berulang dengan cara rekursif untuk setiap subset sampai salah satu dari kondisi berikut ini terpenuhi:
 1. Semua sampel dalam subset memiliki label yang sama
 2. Tidak ada atribut yang tersisa untuk dibagi
 3. Tidak ada data dalam subset.
- d. Pohon Keputusan
Pohon keputusan dibangun dari node dan cabang, di mana setiap node internal menggambarkan pengujian terhadap atribut, sedangkan setiap *leaf node* merepresentasikan hasil klasifikasi dari proses tersebut.

2.3 Chi-Square

Chi-Square merupakan suatu metode seleksi fitur yang berbasis filter. *Chi-Square* adalah salah satu metode yang penyeleksian fitur menggunakan teori statistika yang menguji independensi sebuah term dengan kategorinya [12]. Uji *Chi-Square* merupakan salah satu metode statistik non-parametrik yang digunakan untuk menguji hubungan antara dua variabel kategorik atau untuk menguji kesesuaian distribusi data dengan distribusi yang diharapkan. Menurut Hamsir dan Hamria (2023) didalam penelitiannya menjelaskan bahwa algoritma *chi-square* sebagai seleksi fitur terbukti meningkatkan kinerja dari suatu model dalam proses klasifikasi data [13]. Nilai *Chi-Square* yang tinggi menunjukkan suatu fitur memiliki hubungan yang signifikan dengan kelas target. Rumus dari metode Chi-Square adalah sebagai berikut [14].

$$X^2 = \sum_{i=1}^n \sum_{j=1}^m \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \quad (3)$$

Keterangan :

O_{ij} adalah nilai sebenarnya pada baris ke- i dan kolom ke- j

E_{ij} adalah nilai harapan pada baris ke- i dan kolom ke- j .

3. HASIL DAN PEMBAHASAN

3.1 Identifikasi Masalah

Pada tahap ini, penulis melaksanakan proses identifikasi permasalahan yang berkaitan dengan isu – isu data mining mengenai masa studi mahasiswa Universitas Jambi (UNJA). Hal ini bertujuan untuk merumuskan rencana kerja yang sistematis serta mengidentifikasi jenis data yang diperlukan untuk mendukung penelitian ini.

a. Analisis Sistem Yang Sedang Berjalan

Setiap tahun, jumlah mahasiswa baru di Universitas Jambi terus mengalami peningkatan, dan jumlah data mahasiswa yang tersimpan dalam database juga semakin besar. Namun, hingga saat ini, Universitas Jambi masih memiliki potensi yang belum sepenuhnya dioptimalkan dalam memanfaatkan data mahasiswa yang tersimpan di database untuk mendukung penelitian di bidang data mining, sehingga hasil analisisnya dapat lebih terintegrasi dengan pengambilan keputusan strategis.

3.2 Pembahasan

Penelitian ini bertujuan untuk mengidentifikasi pola dan faktor yang mempengaruhi durasi masa studi dengan menggunakan metode *Decision Tree*. Melalui analisis ini, diharapkan dapat diperoleh pemahaman yang lebih mendalam mengenai variabel-variabel penting yang berkontribusi terhadap keberhasilan atau hambatan kelulusan mahasiswa.

a. Profil Data Penelitian

Pada penelitian ini, data yang digunakan merupakan data lulusan atau alumni Fakultas Sains dan Teknologi Universitas Jambi (UNJA) dalam rentang waktu lima tahun, yaitu dari tahun 2019 hingga tahun 2024. Data ini mencakup sebanyak 2437 lulusan yang menjadi dasar untuk analisis dalam penelitian ini. Informasi mengenai data lulusan ini disajikan secara visual pada Gambar 3.

index	id_alumni	no_mhs	nama_mahasiswa	program_studi	nama_jenjang	fakultas	tanggal_lulus	smt1	smt2	smt3	smt4	smt5	smt6	lama_studi_tahun	keterangan
1	53330	F1C114364	RIKKA LESTARI	KOMPA	S1	Sains dan Teknologi	2019-01-23 00:00:00	3.17	3.25	3.22	3.16	3.2	3.2	4.41	Tidak Taper Waktu
2	53337	F1C114364	PUTRI PERMABANDARI	KOMPA	S1	Sains dan Teknologi	2019-01-23 00:00:00	3.4	3.26	3.17	3.07	3.28	3.06	4.41	Tidak Taper Waktu
3	53558	F1C114370	DEVY AMBERINA SAPUTRI	KOMPA	S1	Sains dan Teknologi	2019-01-23 00:00:00	2.9	2.77	2.75	2.63	2.67	2.72	4.41	Tidak Taper Waktu
4	53560	F1C114366	LUCIANA	KOMPA	S1	Sains dan Teknologi	2019-01-23 00:00:00	2.71	2.8	2.9	2.76	2.86	2.83	4.42	Tidak Taper Waktu
5	53562	F1C114364	INTAN KARLINA	KOMPA	S1	Sains dan Teknologi	2019-01-24 00:00:00	3.31	3.25	3.19	2.96	2.97	2.95	4.42	Tidak Taper Waktu
6	53321	F0A015014	YUNWANDA	KOMPA INDUSTRI	D3	Sains dan Teknologi	2019-01-04 00:00:00	3.21	2.96	2.72	2.95	3.08	3.11	3.45	Taper Waktu
7	53341	F1C114332	NEKA NUGOLIA VITRI	KOMPA	S1	Sains dan Teknologi	2019-01-07 00:00:00	3.1	3.16	3.16	3.15	3.2	3.26	4.42	Tidak Taper Waktu
8	53284	F1C114356	SYUHADA FRATINI	KOMPA	S1	Sains dan Teknologi	2019-01-07 00:00:00	3.07	3.08	3.05	2.84	2.93	3.0	4.42	Tidak Taper Waktu
9	53338	F1C114340	IRINDA MARINI	KOMPA	S1	Sains dan Teknologi	2019-01-07 00:00:00	3.36	3.16	3.12	3.07	3.06	3.06	4.42	Tidak Taper Waktu
10	53453	F1C114368	ANDY YANOTIYAMA	TEKNIK PERSTAMBANGAN	S1	Sains dan Teknologi	2019-01-08 00:00:00	3.38	3.48	3.53	3.57	3.56	3.52	4.43	Tidak Taper Waktu
11	53452	F1C114318	NOVIA PIRMAHYANI	TEKNIK GEOFISIKA	S1	Sains dan Teknologi	2019-01-08 00:00:00	3.4	3.51	3.39	3.33	3.28	3.28	4.43	Tidak Taper Waktu
12	53322	F0A015000	IRVANA PUTRA	KOMPA INDUSTRI	D3	Sains dan Teknologi	2019-01-09 00:00:00	2.21	2.48	2.61	2.63	2.68	3.11	3.47	Taper Waktu
13	62878	F0B015021	ANNISA NABILA DARWANA ERZA	ANALIS KOMPA	D3	Sains dan Teknologi	2019-01-10 00:00:00	2.11	2.37	2.63	2.81	3.0	3.04	3.86	Taper Waktu
14	53263	F0B015010	HEDE ANNA PRAMONO	ANALIS KOMPA	D3	Sains dan Teknologi	2019-01-10 00:00:00	2.45	2.57	2.79	2.96	3.04	3.07	3.47	Taper Waktu
15	53221	F0A015016	NAWAL ANANDA YUSUF	KOMPA INDUSTRI	D3	Sains dan Teknologi	2019-01-10 00:00:00	2.24	2.52	2.73	2.96	3.12	3.16	3.47	Taper Waktu
16	53404	F1C114322	YULIANA ASSADAT	TEKNIK PERSTAMBANGAN	S1	Sains dan Teknologi	2019-01-10 00:00:00	2.93	3.2	3.25	3.17	3.2	3.15	4.43	Tidak Taper Waktu
17	53506	F0A015011	RODRI ALI SAPUTRA HASBULAH	KOMPA INDUSTRI	D3	Sains dan Teknologi	2019-01-11 00:00:00	2.21	2.4	2.6	2.67	3.02	3.06	3.47	Taper Waktu
18	53438	F1C114323	DINDA SATEVA LATIFA	TEKNIK GEOLOGI	S1	Sains dan Teknologi	2019-01-14 00:00:00	3.76	3.77	3.7	3.72	3.76	3.76	4.44	Tidak Taper Waktu
19	53367	F0B015036	SAMA INSIAM PUTRA	ANALIS KOMPA	D3	Sains dan Teknologi	2019-01-14 00:00:00	2.65	2.95	2.84	3.12	3.19	3.16	3.48	Taper Waktu
20	53468	F1C114336	ROZA HEDANIK	TEKNIK GEOFISIKA	S1	Sains dan Teknologi	2019-01-15 00:00:00	3.52	3.49	3.3	3.27	3.24	3.23	4.45	Tidak Taper Waktu
21	53459	F1C114322	FARISA RANDIYAMA	TEKNIK GEOFISIKA	S1	Sains dan Teknologi	2019-01-15 00:00:00	3.62	3.56	3.66	3.65	3.64	3.64	4.45	Tidak Taper Waktu
22	53279	F0B015032	NANDA FERDINANTY	ANALIS KOMPA	D3	Sains dan Teknologi	2019-01-16 00:00:00	2.19	2.32	2.76	2.99	2.92	3.47	Taper Waktu	
23	53457	F1C114312	MERTANTI KUSUMA WINDANI	TEKNIK PERSTAMBANGAN	S1	Sains dan Teknologi	2019-01-17 00:00:00	3.26	3.42	3.45	3.54	3.47	3.42	4.45	Tidak Taper Waktu
24	53266	F0A015016	M. SACRIAN	KOMPA INDUSTRI	D3	Sains dan Teknologi	2019-01-17 00:00:00	2.26	2.45	2.66	2.86	3.12	3.0	3.49	Taper Waktu
25	53342	F1C114302	SUSI HANIKHANI	KOMPA	S1	Sains dan Teknologi	2019-01-17 00:00:00	3.36	3.37	3.3	3.25	3.3	3.26	4.45	Tidak Taper Waktu

Gambar 3. Data Lulusan Fakultas Sains dan Teknologi Universitas Jambi

Data ini mencakup berbagai atribut penting yang relevan untuk penelitian, seperti program studi, tanggal lulus, angkatan, indeks prestasi kumulatif (IPK) dari semester 1 hingga semester 6, dan status kelulusan. grafik berikut menyajikan data jumlah lulusan dari tahun 2019 hingga 2024.

Penelitian ini tidak hanya berfokus pada analisis data tetapi juga bertujuan untuk menghasilkan model klasifikasi yang dapat membantu institusi pendidikan dalam merumuskan strategi peningkatan kualitas lulusan, sehingga universitas dapat lebih proaktif dalam memberikan intervensi atau bimbingan kepada mahasiswa selama masa studi mereka, sehingga dapat meningkatkan peluang keberhasilan dalam menyelesaikan masa studi mahasiswa.

b. Preprocessing Data

Pada tahap ini dilakukan proses pembersihan data dari noise, mengidentifikasi atau menghapus outlier dan menyelesaikan inkonsistensi data menggunakan bahasa python3 dengan menggunakan tools Google Collaboration. Pada proses pembersihan menggunakan google collaboration data terlebih dahulu disajikan dalam bentuk dataframe dengan memanfaatkan library pandas pada python3, berikut proses penyajian data dalam frame pada gambar 4.

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt

# importing the dataset by reading the csv file
data = pd.read_excel('dataAlumni.xlsx', skiprows=1)

# create dataframe
df = pd.DataFrame(data)

# Mengubah indeks baris agar dimulai dari 1
df.index = df.index + 1
df
```

Gambar 4. Library penyajian data dalam frame

Setelah itu, dilakukan proses pengecekan noise pada setiap atribut data untuk memastikan apakah pada data tersebut terdapat nilai null atau kosong pada setiap atribut dengan hasil tidak ada data yang bersifat null, dapat dilihat pada gambar 5.

```
df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2437 entries, 1 to 2437
Data columns (total 15 columns):
 #   Column                Non-Null Count  Dtype
---  ---
 0   id_alumni             2437 non-null   int64
 1   no_mhs                2437 non-null   object
 2   nama_mahasiswa        2437 non-null   object
 3   program_studi         2437 non-null   object
 4   nama_jenjang          2437 non-null   object
 5   fakultas              2437 non-null   object
 6   tanggal_lulus         2437 non-null   datetime64[ns]
 7   smt1                  2437 non-null   float64
 8   smt2                  2437 non-null   float64
 9   smt3                  2437 non-null   float64
10  smt4                  2437 non-null   float64
11  smt5                  2437 non-null   float64
12  smt6                  2437 non-null   float64
13  lama_studi_tahun      2437 non-null   float64
14  keterangan            2437 non-null   object
dtypes: datetime64[ns](1), float64(7), int64(1), object(6)
memory usage: 285.7+ KB
```

Gambar 5 Hasil Pengecekan Data Noise

Selain dilakukan proses pengecekan dan pembersihan data dari data yang bernilai null atau kosong, juga dilakukan pengecekan dan pembersihan data dari redundansi atau data duplikat, dan didapatkan hasil dari pengecekan tidak terdapat data yang bersifat redundan, hasil dari pengecekan data dari redundansi dapat dilihat pada gambar 6.

```
#Check duplicated value
df.duplicated().sum()

0
```

Gambar 6. Hasil Pengecekan Data Duplikat

Setelah dilakukan pengecekan dan pembersihan data dari *noise* dan redundansi pada penelitian ini juga dilakukan transformasi data terhadap atribut "keterangan" yang pada awal bernilai 'Tidak Tepat Waktu' dan 'Tepat Waktu' menjadi bernilai angka yang dimana nilai 1 untuk 'Tepat Waktu' dan nilai 2 untuk 'Tidak Tepat Waktu', proses transformasi data dapat dilihat pada gambar 7.

```
[ ] df_jenis_keterangan = df.copy()
def transformasi_jenis_keterangan(keterangan):
    if keterangan in ['Tepat Waktu']:
        return "1"
    else:
        return "2"
df['jenis_keterangan'] = df['keterangan'].apply(transformasi_jenis_keterangan)
```

Gambar 7. Proses Transformasi Data

Berdasarkan proses transformasi data dalam penelitian yang dimana terdapat dua kategori utama dalam pelabelan status kelulusan mahasiswa, yaitu "Tepat Waktu" dan "Tidak Tepat Waktu". Dari total data yang dianalisis, sebanyak 449 alumni atau sekitar 18,4% dari total populasi mahasiswa berhasil menyelesaikan studi mereka tepat waktu. Sementara itu, mayoritas alumni, yaitu 1.988 mahasiswa atau sekitar 81,6%, tergolong dalam kategori "Tidak Tepat Waktu".

Tabel 1. Perbandingan data keterangan/pelabelan

Keterangan	Label Transformasi	Jumlah Data
Lulus Tepat Waktu	1	449
Lulus Tidak Tepat Waktu	2	1988

Perbandingan ini menunjukkan bahwa jumlah mahasiswa yang lulus tepat waktu jauh lebih rendah dibandingkan dengan mereka yang mengalami keterlambatan dalam menyelesaikan studi. Faktor-faktor yang memengaruhi keterlambatan ini dapat mencakup berbagai aspek, seperti kesulitan akademik, keterlibatan dalam aktivitas ekstrakurikuler, atau faktor sosial-ekonomi. Hasil ini menjadi indikator penting bagi institusi pendidikan dalam merancang strategi peningkatan kualitas akademik dan bimbingan mahasiswa untuk meningkatkan persentase kelulusan tepat waktu.

c. Pembagian Data *Training* dan Data *Testing*

Pada tahap ini, dilakukan proses pembagian data *training* untuk melakukan pemodelan klasifikasi dan data *testing* untuk evaluasi model. Proses pembagian data dilakukan dibuat dalam proses data mining dengan menggunakan *library train_test_split* pada *python3* dengan perbandingan data 70% untuk data *training* dan 30% untuk data *testing*. Setelah dilakukan proses pembagian data menggunakan *library train_test_split* dalam Python, diperoleh data *training* sebanyak 1705 data dan data *testing* sebanyak 732 data. Pembagian ini dilakukan dengan tujuan agar model dapat belajar dari data *training* dan diuji menggunakan data *testing* untuk mengevaluasi performanya. Rincian pembagian data dapat dilihat pada tabel 2.

Tabel 2. Hasil Pembagian *Data Training dan Testing*

No	Keterangan	Jumlah Data
1	Data <i>Training</i>	1705
2	Data <i>Testing</i>	732

Berdasarkan hasil pembagian data pada tabel di atas, proporsi data *training* dan *testing* yang digunakan adalah sekitar 70% untuk pelatihan dan 30% untuk pengujian. Proporsi ini cukup umum digunakan dalam pemodelan machine learning karena memberikan keseimbangan antara jumlah data untuk pembelajaran model dan evaluasi performanya. Dengan demikian, model memiliki cukup informasi untuk belajar tanpa kehilangan kemampuan generalisasi terhadap data baru.

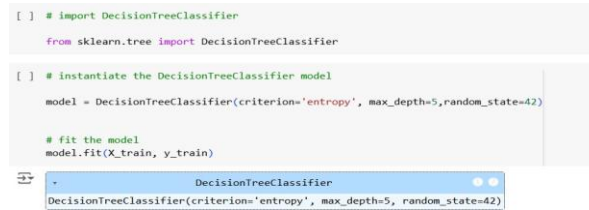
d. Pemodelan *Decision Tree* Tanpa *Feature Selection*

Pada proses pengolahan data ini, algoritma *Decision Tree* digunakan untuk membangun model klasifikasi yang bertujuan untuk mengelompokkan data ke dalam kelas-kelas tertentu berdasarkan fitur yang tersedia. Model dikembangkan menggunakan pustaka *sklearn.tree* dengan memanfaatkan kelas *DecisionTreeClassifier*. Parameter yang digunakan meliputi kriteria pemilihan atribut *entropy*, yang mendasarkan pemilihan atribut terbaik pada perhitungan *Entropy* dan *Information Gain*. Selain itu, kedalaman maksimum pohon keputusan ditetapkan pada nilai 5 (*max_depth=5*) untuk menghindari *overfitting*, dan parameter *random_state = 42* digunakan untuk memastikan hasil pengujian yang konsisten, sesuai pada gambar 8.

```
[ ] # import DecisionTreeClassifier
from sklearn.tree import DecisionTreeClassifier

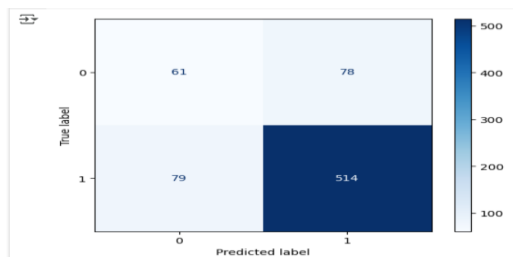
[ ] # instantiate the DecisionTreeClassifier model
model = DecisionTreeClassifier(criterion='entropy', max_depth=5, random_state=42)

# fit the model
model.fit(X_train, y_train)
```



Gambar 8. Proses Pemodelan *Decision Tree* tanpa *Feature Selection*

Model dilatih menggunakan data pelatihan (X_{train} dan y_{train}) melalui fungsi $fit()$, yang menghasilkan struktur pohon keputusan berdasarkan pola dalam data pelatihan. Setelah model selesai dilatih, data uji (X_{test}) diklasifikasikan menggunakan fungsi $predict()$, menghasilkan prediksi yang kemudian dibandingkan dengan label sebenarnya untuk mengevaluasi kinerja model. Evaluasi ini dilakukan dengan *confusion matrix*, yang memberikan gambaran tentang jumlah prediksi benar dan salah yang dihasilkan oleh model dalam setiap kelas.



Gambar 9. Hasil *Confusion Matrix* Pemodelan *Decision Tree* tanpa *Feature Selection*

Berdasarkan hasil gambar 9, model telah diuji untuk mengklasifikasikan data ke dalam dua kelas, yaitu kelas negatif (0) dan kelas positif (1). Hasil evaluasi menunjukkan bahwa model memiliki performa yang cukup baik, meskipun masih terdapat beberapa kesalahan dalam prediksi. Dari confusion matrix, dapat dilihat bahwa model berhasil mengklasifikasikan 514 data positif dengan benar (*True Positive/TP*) dan 61 data negatif dengan benar (*True Negative/TN*). Namun, model juga menghasilkan 78 kesalahan prediksi positif (*False Positive/FP*), di mana data yang sebenarnya negatif diklasifikasikan sebagai positif. Selain itu, terdapat 79 kesalahan prediksi negatif (*False Negative/FN*), di mana data yang sebenarnya positif justru diklasifikasikan sebagai negatif. Dari proses *confusion matrix* diperoleh nilai evaluasi dengan tingkat akurasi, presisi, *recall* dan *F1-Score* sebagai berikut.

Tabel 3. Hasil Evaluasi Kinerja Model Tanpa Seleksi Fitur

Matriks Evaluasi	Nilai
Akurasi	78,57%
Presisi	86,82%
<i>Recall</i>	86,72%
<i>F1-Score</i>	86,77%

Berdasarkan hasil analisis evaluasi kinerja model menggunakan *confusion matrix* secara keseluruhan, algoritma *Decision Tree* memberikan hasil klasifikasi yang cukup baik dan memiliki keunggulan utama pada kemampuannya untuk menghasilkan model yang mudah diinterpretasikan.

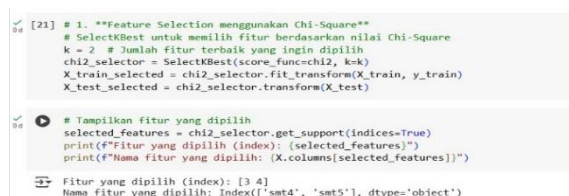
e. Pemodelan *Decision Tree* Dengan *Feature Selection* (*Chi-Square*)

Pada tahap ini dilakukan proses pemodelan klasifikasi data diawali dengan seleksi fitur menggunakan metode *Chi-Square* untuk memilih fitur-fitur terbaik yang berkontribusi signifikan terhadap target klasifikasi. Dalam tahap ini, library *SelectKBest* dari *sklearn.feature_selection* diterapkan dengan parameter $k=2$, yang menentukan jumlah fitur terbaik yang dipilih. Hasil seleksi menunjukkan bahwa fitur yang dipilih adalah *smt4* dan *smt5*, yang menjadi dasar untuk membangun model klasifikasi, dapat dilihat pada gambar 10.

```
[21] # 1. **Feature Selection menggunakan Chi-Square**
# SelectKBest untuk memilih fitur berdasarkan nilai Chi-Square
k = 2 # Jumlah fitur terbaik yang ingin dipilih
chi2_selector = SelectKBest(score_func=chi2, k=k)
X_train_selected = chi2_selector.fit_transform(X_train, y_train)
X_test_selected = chi2_selector.transform(X_test)

# Tampilkan fitur yang dipilih
selected_features = chi2_selector.get_support(indices=True)
print(f"Fitur yang dipilih (index): {selected_features}")
print(f>Nama fitur yang dipilih: {X.columns[selected_features]})

Fitur yang dipilih (index): [3 4]
Nama fitur yang dipilih: Index(['smt4', 'smt5'], dtype='object')
```



Gambar 10. Proses Seleksi Fitur dengan Metode *Chi-Square*

Setelah seleksi fitur, model klasifikasi *Decision Tree* dibangun menggunakan pustaka *DecisionTreeClassifier* dari *sklearn.tree*. Kriteria pemilihan atribut ditetapkan pada *entropy*, dengan kedalaman maksimum pohon keputusan dibatasi hingga 5 untuk menghindari *overfitting*. Model dilatih menggunakan data pelatihan yang telah melalui proses seleksi fitur ($X_{train_selected}$) dan target label (y_{train}) pada gambar 11.

```

# 2. **Modeling dengan DecisionTreeClassifier**
# Buat model Decision Tree
from sklearn.tree import DecisionTreeClassifier, plot_tree
model = DecisionTreeClassifier(criterion='entropy', max_depth=5, random_state=42)

# fit the model
model.fit(X_train_selected, y_train)
    
```

Gambar 11. Proses Pemodelan Decision Tree dengan Seleksi Fitur

Setelah proses pelatihan selesai, model diterapkan untuk melakukan klasifikasi terhadap data uji yang telah melalui tahap seleksi fitur ($X_{test_selected}$). Evaluasi ini dilakukan dengan *confusion matrix*, yang memberikan gambaran tentang jumlah prediksi benar dan salah yang dihasilkan oleh model dalam setiap kelas. Model klasifikasi telah diuji menggunakan dataset uji dan dievaluasi menggunakan *confusion matrix* dengan menghasilkan model yang berhasil mengklasifikasikan 38 sampel dengan benar sebagai negatif (*True Negative* - TN) dan 547 sampel sebagai positif (*True Positive* - TP). Namun, terdapat 101 sampel yang seharusnya negatif tetapi diklasifikasikan sebagai positif (*False Positive* - FP) dan 46 sampel yang seharusnya positif tetapi diklasifikasikan sebagai negatif (*False Negative* - FN). Dari proses *confusion matrix* diperoleh nilai evaluasi dengan tingkat akurasi, presisi, *recall* dan *F1-Score* sebagai berikut.

Tabel 4. Hasil Evaluasi Kinerja Model dengan seleksi fitur

Matriks Evaluasi	Nilai
Akurasi	78,57%
Presisi	86,82%
<i>Recall</i>	86,72%
F1-Score	86,77%

Ketika seleksi fitur diterapkan, terdapat perubahan pada metrik evaluasi model. Akurasi meningkat menjadi 80.00%, yang menunjukkan sedikit peningkatan dalam kemampuan model untuk membuat prediksi yang benar secara keseluruhan. Namun, presisi mengalami sedikit penurunan menjadi 84,41%, sedangkan *recall* meningkat cukup signifikan menjadi 92,24%. Peningkatan *recall* ini menunjukkan bahwa model menjadi lebih sensitif dalam mengenali kelas positif, meskipun dengan sedikit kompromi pada presisi. F1-score juga meningkat menjadi 88,13%, mencerminkan keseimbangan yang lebih baik antara presisi dan *recall* dibandingkan dengan model tanpa seleksi fitur.

Perbedaan hasil ini menunjukkan bahwa seleksi fitur memengaruhi performa model dalam aspek yang berbeda. Tanpa seleksi fitur, model memiliki presisi yang lebih tinggi, yang berarti lebih sedikit prediksi positif yang salah. Namun, dengan seleksi fitur, *recall* meningkat, yang berarti lebih banyak sampel positif yang berhasil terdeteksi. Peningkatan F1-score dengan seleksi fitur juga menunjukkan bahwa model tetap menjaga keseimbangan antara presisi dan *recall*. Analisis ini menunjukkan bahwa pendekatan seleksi fitur dapat membawa perubahan dalam karakteristik performa model, tergantung pada kebutuhan spesifik dari klasifikasi yang diinginkan.

4. KESIMPULAN

Berdasarkan hasil penelitian yang dilakukan terkait klasifikasi data kelulusan mahasiswa Universitas Jambi dengan memanfaatkan metode *Decision Tree* serta teknik seleksi fitur *Chi-Square*, diperoleh temuan bahwa proses seleksi fitur memberikan kontribusi signifikan terhadap peningkatan kinerja model klasifikasi. Sebelum dilakukan seleksi fitur, akurasi model *Decision Tree* tercatat sebesar 78,57%. Namun, setelah fitur-fitur yang paling relevan dipilih melalui metode *Chi-Square*, akurasi meningkat menjadi 79,92%. Peningkatan kinerja model tidak hanya terlihat dari aspek akurasi, tetapi juga tercermin dari metrik evaluasi lainnya, seperti *recall* dan *F1-score*. Nilai *recall* mengalami peningkatan dari 86,72% menjadi 92,24%, yang mengindikasikan bahwa model lebih mampu mengenali kasus-kasus kelulusan secara benar setelah dilakukan seleksi fitur. Demikian pula, nilai *F1-score* juga mengalami peningkatan dari 86,77% menjadi 88,13%, yang menunjukkan keseimbangan yang lebih baik antara presisi dan *recall* dalam model hasil seleksi fitur. Meskipun terdapat sedikit penurunan pada nilai presisi dari

86,82% menjadi 84,41%, hal tersebut tidak secara signifikan mengurangi performa keseluruhan model. Secara keseluruhan, hasil ini menunjukkan bahwa penerapan seleksi fitur dapat membantu mengurangi kompleksitas data tanpa mengorbankan kualitas klasifikasi, bahkan justru meningkatkan kemampuan generalisasi model. Selain itu, seleksi fitur juga berperan penting dalam mengeliminasi atribut-atribut yang tidak relevan atau redundan, sehingga model dapat berfokus pada variabel-variabel yang paling berpengaruh terhadap prediksi kelulusan mahasiswa. Dengan demikian, penggunaan *Chi-Square* sebagai metode seleksi fitur dalam pemodelan klasifikasi berbasis Decision Tree terbukti efektif dalam meningkatkan akurasi dan efisiensi proses klasifikasi.

REFERENCES

- [1] B. D. Ginting, L. Arliana, and N. Kadim, "Penerapan Algoritma K-Nearest Neighbor untuk Klasifikasi Usaha Masyarakat Berdasarkan Jenis Izin Usaha," vol. 2, no. 4, 2024.
- [2] A. I. Saputra, H. Oktavianto, H. Azizah, and A. Faruq, "Penerapan Algoritma Modified K-Nearest Neighbour (MKNN) Pada Klasifikasi Masa Studi Mahasiswa Teknik Informatika Application of the Modified K-Nearest Neighbour (MKNN) Algorithm In the Classification of the Study Period of Informatics Engineering Student," *J. Smart Teknol.*, vol. 3, no. 1, pp. 2774–1702, 2021, [Online]. Available: <http://jurnal.unmuhjember.ac.id/index.php/JST>
- [3] M. R. Qisthiano, P. A. Prayesy, and I. Ruswita, "Penerapan Algoritma Decision Tree dalam Klasifikasi Data Prediksi Kelulusan Mahasiswa," *G-Tech J. Teknol. Terap.*, vol. 7, no. 1, pp. 21–28, 2023, doi: 10.33379/gtech.v7i1.1850.
- [4] N. Nailil Amani, M. Martanto, and U. Hayati, "Penggunaan Algoritma Decision Tree Untuk Prediksi Prestasi Siswa Di Sekolah Dasar Negeri 3 Bayalangu Kidul," *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 8, no. 1, pp. 473–479, 2024, doi: 10.36040/jati.v8i1.8355.
- [5] E. Karyadiputra and Z. Zaenuddin, "Penerapan Algoritma Decision Tree C4.5 Berbasis Seleksi Atribut Chi Squared Untuk Klasifikasi Tingkat Pengetahuan Ibu Dalam Pemberian Asi Eksklusif Pada Bayi," *Technol. J. Ilm.*, vol. 11, no. 1, p. 7, 2020, doi: 10.31602/tji.v11i1.2685.
- [6] O. P. Moerdyanto and I. K. D. Nuryana, "Prediksi Kelulusan Tepat Waktu Menggunakan Pendekatan Pohon Keputusan Algoritma Decision Tree," *J. Informatics Comput. Sci.*, vol. 05, no. 1, pp. 90–96, 2023, [Online]. Available: <https://ejournal.unesa.ac.id/index.php/jinacs/article/view/55329>
- [7] A. Saputra, T. Azhima, and Y. Siswa, "DALAM MODEL KASUS PREDIKSI KETERLAMBATAN BIAYA KULIAH," *J. Inform. dan Komputer*, vol. 6, no. 2, pp. 231–241, 2022.
- [8] M. Darmaputra and D. S. Nugroho, "STUDENT GRADUATION PREDICTION USING ALGORITMA K- MEANS WITH FITUR SELECTION CHI SQUARE," *Equilibria Pendidikan*, vol. 6, no. 2, pp. 45–48, 2017, [Online]. Available: <http://journal.upgris.ac.id/index.php/equilibriapendidikan>
- [9] V. R. Joseph, "Optimal ratio for data splitting," *Stat. Anal. Data Min.*, vol. 15, no. 4, pp. 531–538, 2022, doi: 10.1002/sam.11583.
- [10] Sriani, Aidil Halim Lubis, and Sofiah, "Classification of Patient Satisfaction Level on Health Services Using the C4.5 Algorithm," *J. Ris. Inform.*, vol. 6, no. 2, pp. 95–102, 2024, doi: 10.34288/jri.v6i2.283.
- [11] E. Andriyani, W., *Matematika Pada Kecerdasan Buatan*. CV Tohar Media, 2019.
- [12] T. Ernayanti, M. Mustafid, A. Rusgiyono, and A. R. Hakim, "Penggunaan Seleksi Fitur Chi-Square Dan Algoritma Multinomial Naïve Bayes Untuk Analisis Sentimen Pelanggan Tokopedia," *J. Gaussian*, vol. 11, no. 4, pp. 562–571, 2023, doi: 10.14710/j.gauss.11.4.562-571.
- [13] H. Hamria and H. Hamsir, "K-Nearest Neighbor Berbasis Seleksi Atribut Chi Square Untuk Klasifikasi Penerima Beasiswa," *Simetris J. Tek. Mesin, Elektro dan Ilmu Komput.*, vol. 14, no. 1, pp. 1–10, 2023, doi: 10.24176/simet.v14i1.9178.
- [14] H. M. Lumbantobing, R. A. Marcellino, and I. C. Bu'ulolo, "Penerapan Metode Feature Selection pada Algoritma Naïve Bayes dalam Kasus Keyword Extraction," *Citee*, pp. 117–123, 2020.
- [15] H. Rifa'i, Ryan Hamonangan, Dian Ade Kurnia, Kaslani, and Mulyawan, "Implementasi Algoritma Decision Tree Dalam Klasifikasi Kompetensi Siswa," *KOPERTIP J. Ilm. Manaj. Inform. dan Komput.*, vol. 6, no. 1, pp. 15–20, 2022, doi: 10.32485/kopertip.v6i1.131.
- [16] Endang Etriyanti, "Perbandingan Tingkat Akurasi Metode Knn Dan Decision Tree Dalam Memprediksi Lama Studi Mahasiswa," *J. Ilm. Bin. STMIK Bina Nusantara. Jaya Lubuklinggau*, vol. 3, no. 1, pp. 6–14, 2021, doi: 10.52303/jb.v3i1.40.
- [17] Y. Crismayella, N. Satyahadewi, and H. Perdana, "Algoritma Adaboost pada Metode Decision Tree untuk Klasifikasi Kelulusan Mahasiswa," *Jambura J. Math.*, vol. 5, no. 2, pp. 278–288, 2023, doi: 10.34312/jjom.v5i2.18790.
- [18] C. N. Dengen, K. Kusriani, and E. T. Luthfi, "Implementasi Decision Tree Untuk Prediksi Kelulusan Mahasiswa Tepat aktu," *Sisfotenika*, vol. 10, no. 1, p. 1, 2020, doi: 10.30700/jst.v10i1.484.