Penerapan Multinomial Naïve Bayes dan Seleksi Fitur Gain Ratio dalam Memprediksi Topik Skripsi Berbasis Rekam Jejak Akademik

Omanius Seda Oja[1], Rahmatina Hidayati[2]
Sistem Informasi, Fakultas Teknologi Informasi, Universitas Merdeka Malang
e-mail: sedaoja@gmail.com[1], rahmatina.hidayati@unmer.ac.id[2]

Kata Kunci:

Prediksi Topik Skripsi
Multinomial Naïve Bayes
Gain Ratio
Rekam Jejak Akademik

ABSTRAK

Penentuan topik skripsi menjadi salah satu kendala yang dihadapi mahasiswa karena pilihan bidang penelitian cukup luas dan belum selalu didasarkan pada kecenderungan kompetensi akademik yang terukur. Penelitian ini bertujuan membangun model prediksi dan rekomendasi kategori topik skripsi mahasiswa menggunakan Multinomial Naïve Bayes berdasarkan rekam jejak nilai mata kuliah serta menganalisis pengaruh seleksi fitur Gain Ratio. Data penelitian terdiri atas 256 data mahasiswa alumni Program Studi Sistem Informasi dengan 27 mata kuliah inti sebagai atribut prediktor dan lima kategori topik skripsi sebagai kelas target, yaitu Data Mining dan AI, Jaringan dan Keamanan, Sistem Informasi, UI/UX, serta Web dan Mobile. Hasil pengujian menunjukkan Skenario 1 dengan Gain Ratio memiliki performa lebih tinggi, yaitu accuracy 74,23%, precision 76,05%, recall 73,88%, dan F1-Macro 73,77%. Skenario 2 tanpa Gain Ratio memperoleh accuracy 72,68% dan F1-Macro 70,21%. Meskipun demikian, Skenario 2 dipilih sebagai model akhir karena mempertahankan seluruh 27 mata kuliah. Implementasi pada 100 mahasiswa angkatan 2023 menghasilkan rekomendasi terbanyak pada topik Web dan Mobile sebanyak 47 mahasiswa.

Keyword:

Thesis Topic Prediction
Multinomial Naïve Bayes
Gain Ratio
Academic Records

ABSTRACT

Determining a thesis topic is a challenge for students because the available research fields are broad and the selection process is not always based on measurable academic competencies. This study aims to develop a prediction and recommendation model for student thesis topic categories using Multinomial Naïve Bayes based on course-grade records and to analyze the effect of Gain Ratio feature selection. The dataset consists of 256 alumni records from the Information Systems Study Program, using 27 core courses as predictor attributes and five thesis-topic classes: Data Mining and AI, Network and Security, Information Systems, UI/UX, and Web and Mobile. The evaluation results show that Scenario 1 with Gain Ratio achieved higher performance, with 74.23% accuracy, 76.05% precision, 73.88% recall, and 73.77% macro F1-score. Scenario 2 without Gain Ratio achieved 72.68% accuracy and 70.21% macro F1-score. Nevertheless, Scenario 2 was selected as the final model because it retained all 27 courses. Implementation on 100 active students from the 2023 cohort produced the highest number of recommendations for Web and Mobile, with 47 student

Copyright © SENASIF 2026
ISSN: 2598-0076

I. PENDAHULUAN

Skripsi merupakan salah satu tahap penting dalam penyelesaian pendidikan tinggi karena menjadi bentuk integrasi pengetahuan, keterampilan analitis, dan pemahaman metodologis yang diperoleh mahasiswa selama perkuliahan. Pada Program Studi Sistem Informasi Universitas Merdeka Malang, mahasiswa menempuh mata kuliah yang mencerminkan berbagai kompetensi, mulai dari pemrograman, pengolahan data, sistem informasi, jaringan dan keamanan, hingga interaksi manusia dan komputer. Rekam jejak nilai pada mata kuliah tersebut dapat digunakan untuk menggambarkan kecenderungan kemampuan akademik mahasiswa pada bidang tertentu [1], [2].

Permasalahan muncul ketika mahasiswa harus menentukan topik skripsi yang sesuai dengan kemampuan dan minatnya. Pemilihan topik masih dapat dipengaruhi kebiasaan, tren, maupun saran dari lingkungan sekitar, sehingga belum sepenuhnya berbasis pada data akademik yang terukur. Penelitian terdahulu menunjukkan bahwa penentuan topik skripsi dapat dibantu dengan pendekatan berbasis data. Farid et al. [3] membangun sistem pendukung keputusan rekomendasi topik skripsi menggunakan Naïve Bayes dan memperoleh akurasi 69,27%. Malja et al. [4] juga menunjukkan bahwa Naïve Bayes dapat digunakan untuk mengklasifikasikan bidang minat mahasiswa dalam menentukan topik tugas akhir. Penelitian lain menunjukkan penerapan Naïve Bayes pada data akademik dapat memberikan prediksi yang cukup baik [5],[6].

Walaupun penelitian terdahulu telah menunjukkan potensi klasifikasi berbasis data akademik, masih terdapat ruang untuk meningkatkan relevansi atribut yang digunakan. Dataset akademik dapat memiliki banyak mata kuliah yang tidak semuanya memberikan kontribusi yang sama terhadap kelas target. Seleksi fitur dapat membantu mengurangi dimensi dan mempertahankan atribut yang paling informatif. Gain Ratio merupakan metode seleksi fitur berbasis informasi yang dapat mengurangi bias Information Gain terhadap atribut dengan banyak nilai [7]. Suprapto [8] menunjukkan bahwa kombinasi Naïve Bayes dan metode seleksi fitur, termasuk Gain Ratio, dapat meningkatkan performa pada dataset berdimensi tinggi.

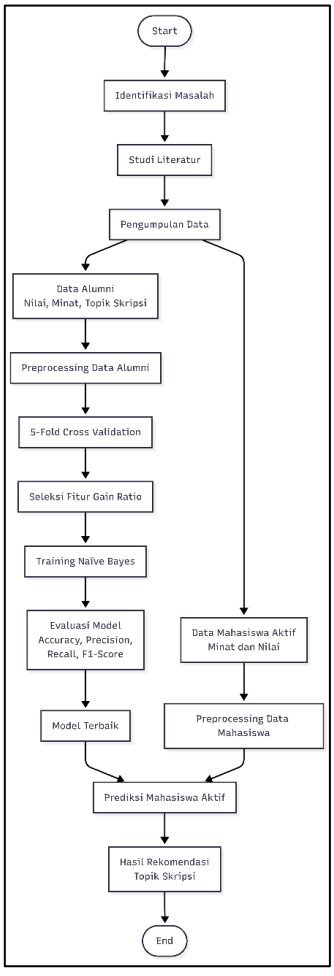
Penelitian ini mengembangkan model prediksi kategori topik skripsi berdasarkan rekam jejak nilai mata kuliah mahasiswa. Lima kategori yang digunakan adalah Data Mining dan AI, Jaringan dan Keamanan, Sistem Informasi, UI/UX, serta Web dan Mobile. Model dibangun menggunakan Multinomial Naïve Bayes dan dievaluasi melalui 5-Fold Cross Validation. Untuk mengatasi ketidakseimbangan kelas, SMOTE diterapkan hanya pada data pelatihan setiap fold. Dua skenario dibandingkan, yaitu Multinomial Naïve Bayes dengan seleksi fitur Gain Ratio dan Multinomial Naïve Bayes tanpa seleksi fitur. Tujuan penelitian adalah mengetahui kinerja kedua skenario serta menentukan model yang paling sesuai untuk menghasilkan rekomendasi topik skripsi bagi mahasiswa aktif.

Copyright © SENASIF 2026
ISSN: 2598-0076

II. METODE

Penelitian ini merupakan penelitian kuantitatif dengan pendekatan eksperimen komputasi pada bidang machine learning. Data yang digunakan berupa data akademik mahasiswa alumni Program Studi Sistem Informasi dan data mahasiswa aktif angkatan 2023. Data alumni digunakan sebagai data pembentukan dan pengujian model, sedangkan data mahasiswa aktif digunakan untuk implementasi prediksi. Dataset alumni yang telah melalui proses validasi terdiri atas 256 data dengan atribut nilai mata kuliah dan judul skripsi sebagai sumber pembentukan kelas target [9], [10].

Alur penelitian dirancang secara bertahap mulai dari identifikasi masalah hingga menghasilkan rekomendasi topik skripsi. Data alumni digunakan sebagai data berlabel untuk proses pembelajaran model, sedangkan data mahasiswa aktif digunakan sebagai data baru untuk memperoleh prediksi. Tahapan utama meliputi pengumpulan dan preprocessing data, pembagian data menggunakan Stratified 5-Fold Cross Validation, penanganan ketidakseimbangan data pada data latih, seleksi fitur Gain Ratio pada Skenario 1, klasifikasi menggunakan Multinomial Naïve Bayes, evaluasi model, pemilihan model, dan implementasi prediksi pada mahasiswa aktif. Pada tahap akhir, kelas dengan probabilitas tertinggi digunakan sebagai rekomendasi topik skripsi. Alur tersebut ditunjukkan pada gambar 1.



Gambar 1. Alur Penelitian

Copyright © SENASIF 2026

ISSN: 2598-0076

Tahap *preprocessing* dilakukan secara berurutan. Pertama, cleaning digunakan untuk memeriksa missing value, duplikasi, inkonsistensi penulisan, dan format data. Kedua, filtering dilakukan untuk memilih mata kuliah yang relevan. Dataset awal memiliki 53 mata kuliah dan disaring menjadi 27 mata kuliah inti dengan menghilangkan mata kuliah dasar umum yang tidak dianggap mewakili kompetensi inti Program Studi Sistem Informasi. Ketiga, nilai huruf didiskritisasi menjadi bentuk numerik agar dapat diproses oleh model. Keempat, judul skripsi diberi label berdasarkan kategori topik. Pelabelan mempertimbangkan kata kunci dan fokus utama judul, kemudian divalidasi secara manual ketika terdapat ambiguitas. Kelima, mata kuliah dikelompokkan berdasarkan bidang kompetensi seperti Data, Sistem Informasi, Pemrograman dan Teknologi Informasi, serta Jaringan dan Keamanan [11], [12].

Distribusi kelas setelah labeling terdiri atas 74 data Sistem Informasi, 67 data Web dan Mobile, 48 data UI/UX, 48 data Data Mining dan AI, serta 19 data Jaringan dan Keamanan. Distribusi tersebut menunjukkan adanya ketidakseimbangan kelas, khususnya pada kategori Jaringan dan Keamanan. Karena itu, pada setiap iterasi 5-Fold Cross Validation dilakukan oversampling menggunakan Synthetic Minority Over-sampling Technique (SMOTE) hanya terhadap data training. Pendekatan ini digunakan untuk mencegah data sintetis masuk ke data pengujian dan mengurangi risiko data leakage.

Validasi model menggunakan Stratified 5-Fold Cross Validation agar proporsi kelas tetap terjaga pada setiap fold. Pada setiap iterasi, empat fold digunakan sebagai data latih dan satu fold sebagai data uji. Untuk Skenario 1, Gain Ratio dihitung pada data pelatihan untuk menentukan fitur paling informatif. Dari 27 atribut, 15 mata kuliah dengan nilai Gain Ratio tertinggi digunakan pada proses klasifikasi. Skenario 2 menggunakan seluruh 27 atribut tanpa seleksi fitur. Selanjutnya, data training yang telah diseimbangkan digunakan untuk melatih Multinomial Naïve Bayes [13].

Evaluasi dilakukan menggunakan accuracy, precision, recall, dan F1-score macro. Accuracy menunjukkan proporsi prediksi benar secara keseluruhan, sedangkan precision dan recall digunakan untuk melihat kemampuan model pada masing-masing kelas. F1-score macro digunakan sebagai metrik utama karena memberikan bobot yang sama kepada setiap kelas dan lebih representatif pada kondisi data tidak seimbang. Hasil dari lima fold dirata-ratakan untuk memperoleh performa keseluruhan. Model terpilih kemudian diterapkan pada data 100 mahasiswa aktif angkatan 2023 dengan menggunakan 27 nilai mata kuliah sebagai input [14].

Tabel 1. Distribusi Kelas Topik Skripsi Hasil *Labeling*

Topik Skripsi	Jumlah
Data Mining dan AI	48
Jaringan dan Keamanan	19
Sistem Informasi	74
UI/UX	48

Copyright © SENASIF 2026
ISSN: 2598-0076

Topik Skripsi	Jumlah
Web dan Mobile	67
Total	256

III. HASIL DAN PEMBAHASAN

Hasil Preprocessing dan Seleksi Fitur

Hasil preprocessing menghasilkan 256 data yang siap digunakan untuk pemodelan. Filtering mengurangi atribut mata kuliah dari 53 menjadi 27 mata kuliah inti. Diskritisasi mengubah nilai huruf menjadi nilai numerik, sedangkan labeling menghasilkan lima kelas target. Tahapan ini penting karena kualitas data dan konsistensi label menjadi dasar bagi proses klasifikasi. Dalam konteks Educational Data Mining, rekam nilai akademik dapat digunakan untuk mengidentifikasi pola dan kompetensi mahasiswa [1].

Pada Skenario 1, Gain Ratio digunakan untuk mengurutkan relevansi atribut. Lima belas fitur dengan nilai tertinggi dipilih. Fitur dengan nilai Gain Ratio paling tinggi adalah Dasar Basis Data sebesar 0,0621, diikuti Strategi Pengelolaan dan Perolehan SI sebesar 0,0512, Analisis dan Desain Sistem sebesar 0,0349, Implementasi dan Evaluasi SI sebesar 0,0331, dan Interaksi Manusia Komputer sebesar 0,0315. Seleksi tersebut menunjukkan bahwa mata kuliah yang berkaitan dengan basis data, analisis sistem, implementasi sistem, dan interaksi manusia-komputer memberikan informasi yang relatif kuat dalam membedakan kelas topik.

```
def select_features_gain_ratio_manual(X_train_num, y_train, fold, skenario, k=15):
    discretizer = KBinsDiscretizer(n_bins=3, encode='ordinal', strategy='uniform')
    X_disc = discretizer.fit_transform(X_train_num)
    df_gr = calculate_gain_ratio_manual(X_disc, y_train, X_train_num.columns,
                                        judul=f"Entropy(S) Fold {fold}")

    # Visualisasi GR per fold
    plt.figure(figsize=(12,6))
    top15 = df_gr.head(15).sort_values('Gain_Ratio')
    sns.barplot(x='Gain_Ratio', y='Fitur', data=top15, palette='viridis')
    plt.title(f'Top 15 Gain Ratio - {skenario} Fold {fold}')
    plt.xlabel('Gain Ratio')
    plt.tight_layout()
    plt.savefig(f"hasil_tahap1/GR_{safe_sheet_name(skenario)}_Fold{fold}.png")
    plt.show()
    plt.close()

    print("\n" + "="*80)
    print(f"TABEL PERHITUNGAN GAIN RATIO - {skenario} FOLD {fold} - 15 TERBAIK")
    print("="*80)
    print(df_gr.head(15).to_string(index=False))
    print("="*80)
    return df_gr['Fitur'].tolist()[:k], df_gr
```

Gambar 2. Kode Perhitungan Gain Ratio

Copyright © SENASIF 2026
ISSN: 2598-0076

Tabel 2. Lima Belas Fitur dengan Gain Ratio Tertinggi

No	Fitur	Information Gain	Split Information	Gain Ratio
1	Dasar Basis Data	0,0928	1,4943	0,0621
2	Strategi Pengelolaan dan Perolehan SI	0,0776	1,5163	0,0512
3	Analisis dan Desain Sistem	0,0471	1,3505	0,0349
4	Implementasi dan Evaluasi SI	0,0504	1,5223	0,0331
5	Interaksi Manusia Komputer	0,0391	1,2404	0,0315
6	Algoritma dan Struktur Data	0,0467	1,5334	0,0305
7	Rekayasa Perangkat Lunak	0,0472	1,5535	0,0304
8	Pemrograman Lanjut	0,0425	1,5215	0,0279
9	Manajemen Risiko dan Audit SI	0,0398	1,4746	0,0270
10	Dasar Pemrograman	0,0346	1,4576	0,0238
11	Sist. Informasi Akuntansi dan Manajemen Bisnis	0,0362	1,5256	0,0237
12	Pengembangan Aplikasi Bisnis Mobile	0,0345	1,5130	0,0228
13	Jaringan Komunikasi Data	0,0325	1,5127	0,0215
14	Statistika Probabilitas	0,0324	1,5454	0,0210
15	Perancangan UI/UX	0,0315	1,5323	0,0206

Evaluasi Model

Hasil 5-Fold Cross Validation pada Skenario 1 menunjukkan performa yang relatif stabil. Accuracy per fold berada pada rentang 71,15% sampai 76,47%, sedangkan F1-Macro berada pada rentang 69,33% sampai 77,48%. Rata-rata accuracy sebesar 74,23%, precision 76,05%, recall 73,88%, dan F1-Macro 73,77%. Hasil ini menunjukkan bahwa penggunaan 15 fitur terpilih dapat membantu model memusatkan proses klasifikasi pada atribut yang lebih informatif [15].

Copyright © SENASIF 2026
ISSN: 2598-0076

Tabel 3. Hasil Evaluasi 5-Fold Skenario 1 MNB dengan Gain Ratio

Fold	*Accuracy*	*Precision*	*Recall*	*F1-Macro*
1	0,7115	0,7187	0,7103	0,7067
2	0,7647	0,8218	0,7568	0,7748
3	0,7647	0,7501	0,8000	0,7572
4	0,7451	0,8126	0,7327	0,7564
5	0,7254	0,6990	0,6940	0,6933
Rata-rata	0,7423	0,7605	0,7388	0,7377

Classification report Skenario 1 memperlihatkan bahwa Data Mining dan AI serta Web dan Mobile memperoleh F1-Score masing-masing 0,79. UI/UX memperoleh 0,74, Jaringan dan Keamanan 0,70, sedangkan Sistem Informasi 0,68. Nilai yang lebih rendah pada Sistem Informasi menunjukkan bahwa pola kelas tersebut masih lebih sulit dibedakan dibandingkan kelas dengan karakteristik fitur yang lebih kuat. Hasil ini sejalan dengan temuan penelitian yang menunjukkan bahwa ketidakseimbangan data dapat memengaruhi kinerja klasifikasi topik skripsi [3].

Skenario 2 menggunakan seluruh 27 mata kuliah tanpa seleksi fitur. Dengan SMOTE, rata-rata accuracy sebesar 72,68%, precision 72,17%, recall 70,40%, dan F1-Macro 70,21%. Dibandingkan kondisi tanpa SMOTE, accuracy meningkat 0,41%, recall meningkat 3,50%, dan F1-Macro meningkat 1,89%, sementara precision turun 2,57%. Temuan ini menunjukkan bahwa SMOTE membantu model mengenali kelas minoritas dengan lebih baik ketika seluruh atribut digunakan.

Tabel 4. Perbandingan Performa Rata-rata Dua Skenario

Skenario	*Accuracy*	*Precision*	*Recall*	*F1-Macro*
Skenario 1 – MNB + Gain Ratio	74,23%	76,05%	73,88%	73,77%
Skenario 2 – MNB tanpa Gain Ratio	72,68%	72,17%	70,40%	70,21%

Pembahasan Pemilihan Model

Secara kuantitatif, Skenario 1 merupakan model dengan performa lebih tinggi. Selisih F1-Macro antara Skenario 1 dan Skenario 2 mencapai 3,56 poin persentase. Hasil tersebut memperkuat argumen bahwa seleksi fitur dapat membantu mengurangi atribut yang kurang informatif dan meningkatkan konsistensi model. Temuan ini juga mendukung penelitian Suprapto [8] mengenai manfaat feature selection pada Naïve Bayes serta penelitian Oktavianti et al. [7] yang menempatkan seleksi berbasis informasi sebagai pendekatan untuk mengurangi dimensi atribut.

Copyright © SENASIF 2026

ISSN: 2598-0076

Namun, penelitian ini memilih Skenario 2 sebagai model akhir untuk implementasi rekomendasi. Pemilihan tersebut tidak didasarkan hanya pada accuracy tertinggi, tetapi pada kebutuhan rekomendasi yang mempertimbangkan seluruh 27 mata kuliah inti. Atribut dengan Gain Ratio rendah tetap dipertahankan karena nilai tersebut tidak otomatis berarti atribut tidak relevan secara kontekstual terhadap suatu topik tertentu. Dengan demikian, Skenario 2 dipandang lebih sesuai untuk tujuan rekomendasi yang komprehensif, meskipun performa klasifikasinya lebih rendah dibandingkan Skenario 1.

Pertimbangan tersebut juga menunjukkan perbedaan antara model untuk evaluasi klasifikasi dan model untuk kebutuhan rekomendasi. Jika tujuan utama adalah memperoleh performa klasifikasi tertinggi, Skenario 1 lebih tepat. Sebaliknya, jika tujuan utama adalah memberikan rekomendasi berdasarkan keseluruhan rekam jejak kurikulum mahasiswa, Skenario 2 memberikan cakupan atribut yang lebih lengkap. Pendekatan ini masih konsisten dengan penelitian sebelumnya yang memanfaatkan nilai mata kuliah sebagai dasar rekomendasi topik skripsi [3], [9].

Implementasi Prediksi Mahasiswa Aktif

Model akhir Skenario 2 diterapkan pada 100 mahasiswa angkatan 2023. Karena mahasiswa tersebut belum memiliki judul skripsi, prediksi dilakukan hanya berdasarkan 27 nilai mata kuliah. Hasilnya menunjukkan empat kategori rekomendasi yang muncul, yaitu Web dan Mobile, Data Mining dan AI, Sistem Informasi, serta Jaringan dan Keamanan. Kategori UI/UX tidak menjadi rekomendasi tertinggi pada 100 data tersebut.

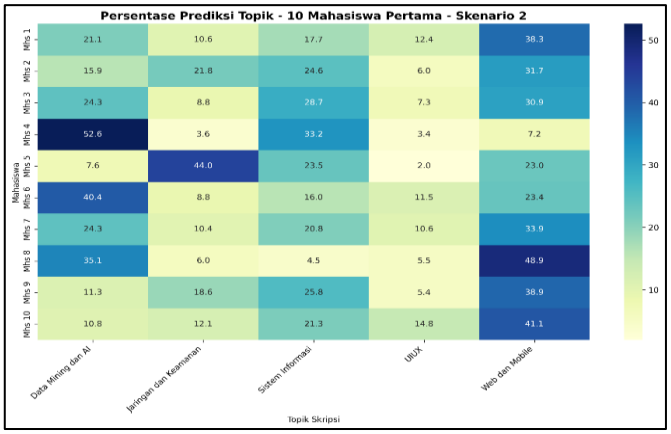
Tabel 5. Distribusi Hasil Prediksi 100 Mahasiswa

Topik Rekomendasi	Jumlah Mahasiswa	Persentase
Web dan Mobile	47	47%
Data Mining dan AI	26	26%
Sistem Informasi	20	20%
Jaringan dan Keamanan	7	7%
UI/UX	0	0%
Total	100	100%

Web dan *Mobile* menjadi rekomendasi paling dominan dengan 47 mahasiswa, diikuti Data Mining dan AI sebanyak 26 mahasiswa, Sistem Informasi sebanyak 20 mahasiswa, dan Jaringan dan Keamanan sebanyak 7 mahasiswa. Selain label prediksi, model menghasilkan probabilitas untuk setiap kategori sehingga rekomendasi tidak hanya berbentuk satu kelas. Sebagai contoh, pada data mahasiswa pertama probabilitas tertinggi berada pada Web dan Mobile sebesar 38,27%, mahasiswa keempat pada Data Mining dan AI sebesar 52,62%, mahasiswa kelima pada Jaringan dan Keamanan sebesar 44,01%, dan mahasiswa kedelapan pada Web dan Mobile sebesar 48,93%. Nilai probabilitas

Copyright © SENASIF 2026
ISSN: 2598-0076

tersebut dapat digunakan sebagai informasi tambahan bagi mahasiswa ketika mempertimbangkan topik yang akan dipilih.



Gambar 3. Persentase Prediksi 10 Mahasiswa

Hasil implementasi memperlihatkan bahwa model dapat digunakan sebagai alat bantu pengambilan keputusan akademik, bukan sebagai penentu tunggal topik skripsi. Rekomendasi tetap perlu dipertimbangkan bersama minat mahasiswa, ketersediaan pembimbing, ruang lingkup penelitian, dan kelayakan topik. Dengan posisi tersebut, model berfungsi memberikan arah awal yang berbasis rekam akademik, sedangkan keputusan akhir tetap berada pada mahasiswa dan pihak akademik.

IV. SIMPULAN

Penelitian ini berhasil menerapkan Multinomial Naïve Bayes untuk memprediksi kategori topik skripsi berdasarkan rekam jejak nilai mata kuliah mahasiswa. Pengolahan data menghasilkan 27 mata kuliah inti sebagai atribut prediktor dan lima kategori topik skripsi sebagai target klasifikasi. Penggunaan 5-Fold Cross Validation dan SMOTE pada data training memberikan mekanisme evaluasi yang lebih terkontrol pada dataset yang tidak seimbang.

Skenario dengan seleksi fitur Gain Ratio menghasilkan performa terbaik secara kuantitatif dengan accuracy 74,23%, precision 76,05%, recall 73,88%, dan F1-Macro 73,77%. Seleksi Gain Ratio mereduksi 27 atribut menjadi 15 atribut dengan nilai informasi tertinggi. Meskipun demikian, Skenario 2 tanpa seleksi fitur dipilih sebagai model akhir karena mempertahankan seluruh 27 mata kuliah dan dinilai lebih sesuai dengan kebutuhan rekomendasi yang komprehensif. Implementasi model pada 100 mahasiswa angkatan 2023 menghasilkan rekomendasi Web dan Mobile untuk 47 mahasiswa, Data Mining dan AI untuk 26 mahasiswa, Sistem Informasi untuk 20 mahasiswa, dan Jaringan dan Keamanan untuk 7 mahasiswa.

Secara keseluruhan, model dapat digunakan sebagai pendukung keputusan awal dalam menentukan arah topik skripsi berdasarkan capaian akademik. Pengembangan berikutnya dapat

Copyright © SENASIF 2026
ISSN: 2598-0076

memperluas jumlah data, menambah variabel non-akademik yang relevan, membandingkan algoritma klasifikasi lain, dan mengembangkan model menjadi sistem rekomendasi yang dapat digunakan secara langsung oleh mahasiswa dan dosen.

DAFTAR RUJUKAN

[1] N. Abdurahman, I. Ripai, W. Baswardono, and H. Maulana, "Sistem Rekomendasi Topik Skripsi Berdasarkan Minat dan Nilai Akademik Mahasiswa Menggunakan Machine Learning di Universitas Muhammadiyah Kuningan dalam Perspektif Rasionalitas Ilmiah," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 10, no. 1, pp. 1406–1414, 2026.

[2] Z. Alamgir, H. Akram, S. Karim, and A. Wali, "Enhancing Student Performance Prediction via Educational Data Mining on Academic Data," vol. 23, no. 1, pp. 1–24, 2024, doi: 10.15388/infedu.2024.04.

[3] Farid, U. Enri, and Y. Umaidah, "Sistem Pendukung Keputusan Rekomendasi Topik Skripsi Menggunakan Naïve Bayes Classifier," *JOINTECS (Journal Inf. Technol. Comput. Sci.*, vol. 6, no. 1, pp. 35–42, 2021, [Online]. Available: http://publishing-widyagama.ac.id/ejournal-v2/index.php/jointecs

[4] F. Malja, D. Kurniadi, K. Budayawan, and R. Darni, "Klasifikasi Bidang Minat Mahasiswa Elektronika Dalam Menentukan Topik Tugas Akhir Menggunakan Algoritma Naïve Bayes Classifier," *J. Pendidik. Tambusai*, vol. 7, pp. 4192–4202, 2023.

[5] D. Sartika, N. Kawadha, P. Gumay, and R. Gustriansyah, "Jurnal Teknologi dan Manajemen Informatika Comparative Analysis of SMOTE , WMOTE , and ADASYN Oversampling Methods on Multinomial Naive Bayes Performance in Classifying Toddlers Nutritional Status," vol. 12, no. 1, pp. 10–20, 2026.

[6] A. Zubair and M. Muksin, "Penerapan Metode Naive Bayes untuk Klasifikasi Status Gizi (Studi Kasus di Klinik Bromo Malang)," in *Seminar Nasional Sistem Informasi*, Fakultas Teknologi Informasi, Universitas Merdeka Malang, 2018, pp. 1204–1208.

[7] E. Oktavianti, M. Agustin, and R. Sari, "Implementasi Algoritma Decision Tree dengan Fitur Seleksi Weight by Information Gain," vol. 9, no. 2, pp. 118–126, 2024.

[8] Suprapto, "Improvement Naive Bayes Menggunakan Forward Selection, Information Gain dan Gain Ratio untuk Penanganan Independensi Fitur," *J. Sos. dan Teknol.*, vol. 5, no. 4, pp. 925–935, 2025.

[9] A. Wijaya, "Sistem pendukung keputusan rekomendasi topik judul skripsi menggunakan metode naive bayes classifier," *JEECOM (Journal Electr. Eng. Comput.*, vol. 4, no. 2, pp. 96–101, 2022.

[10] E. Yumami, M. Jannah, and K. O. Putra, "Klasifikasi Judul Skripsi Mahasiswa Berdasarkan Konsentrasi Program Studi Menggunakan Algoritma Naive Bayes," *J. Comput. Sci. Technol.*, vol. 5, no. 1, pp. 24–28, 2025.

Copyright © SENASIF 2026
ISSN: 2598-0076

[11] R. R. Ekawati, V. Atina, and J. Maulidar, “Prediksi Ketuntasan Belajar Siswa Menggunakan Naive Bayes dengan Integrasi Data Akademik, Absensi, dan Partisipasi Ekstrakurikuler,” *J. Pendidik. dan Teknol. Indones.*, vol. 5, no. 4, pp. 1161–1173, 2025, doi: 10.52436/1.jpti.752.

[12] I. Riadi, R. Umar, and R. Anggara, “Prediksi Kelulusan Tepat Waktu Berdasarkan Riwayat Akademik Menggunakan Metode Naïve Bayes,” *J. Pendidik. Teknol. Informasi,* vol. 4, no. 1, pp. 191–203, 2024.

[13] R. A. Safitri, “Komparasi Metode K-Nearest Neighbor dan Naïve Bayes untuk Mengklasifikasi Resiko Diabetes,” vol. 14, no. 2, pp. 297–303, 2024.

[14] E. M. Stefany, P. R. Ningsih, M. Arif, and L. M. Diana, “Pemetaan Peminatan Penelitian Tugas Akhir Mahasiswa Prodi Pendidikan Informatika Fakultas Ilmu Pendidikan Universitas Trunojoyo Madura,” vol. 9, no. 1, pp. 83–90, 2023.

[15] F. Firdaus and L. P. Pane, “Perbandingan Decision Tree dan Naive Bayes untuk Prediksi Customer Churn,” *JISAD (Journal Inf. Syst. Appl. Dev.*, vol. 4, no. 1, pp. 195–200, 2026, doi: 10.26905/jisad.v4i1.16837.

Copyright © SENASIF 2026
ISSN: 2598-0076