

Segmentasi Kepuasan Mahasiswa Terhadap Dosen Menggunakan K-Means Clustering Dan Identifikasi Faktor Dominan Dengan Random Forest

Restu Rakhmawati¹, Suamanda Ika Novichasari² Imam Adi Nata³ Fadhila Syahida Wibowo⁴, Zharifa Nur Majidah⁵, Meily Adenia⁶

^{1,2,3,4,5,6}Program Studi Teknologi Informasi, Universitas Tidar, Indonesia

¹resturakhma@untidar.ac.id, ²manda@untidar.ac.id, ³imamadinata@untidar.ac.id, ⁴fadhila.syahida.wibowo@students.untidar.ac.id, ⁵zharifanurmajidah@students.untidar.ac.id, ⁶meilyadennia@students.untidar.ac.id

Info Artikel

Riwayat Artikel:

Received 2026-03-10

Revised 2026-04-27

Accepted 2026-05-19

Abstract – This study aims to analyze student satisfaction patterns regarding lecturer teaching performance by integrating K-Means Clustering and Random Forest algorithms. The comprehensive research data includes 5,260 individual observations, which were systematically analyzed based on service quality dimensions and academic attributes. The analytical results, determined by utilizing the Elbow Method, successfully established three optimal clusters. These clusters represent distinct segments: the Very Satisfied segment with a score of 4.84, the Moderately Satisfied segment with a score of 4.04, and the Dissatisfied segment with a score of 3.04. Furthermore, the Random Forest algorithm demonstrated a predictive accuracy of 47% when evaluated on 1,576 test data samples. It successfully identified that the Semester Credit Load (SKS) is the most dominant determinant influencing student satisfaction, with an importance value of 47.46%. A unique finding demonstrates that students with the highest academic workload, averaging 20.81 SKS, are actually positioned within the Very Satisfied segment. Ultimately, this study concludes that a more intensive level of student academic engagement correlates positively with appreciation for lecturer teaching quality. These results provide strategic guidance for university management aiming to improve services based on student academic profiles.

Keywords: Credit Load, K-Means Clustering, Lecturer Performance, Random Forest, Student Satisfaction.

Corresponding Author:

Restu Rakhmawati

Email: resturakhma@untidar.ac.id



This is an open access article under the [CC BY 4.0](https://creativecommons.org/licenses/by/4.0/) license.

Abstrak – Penelitian ini bertujuan menganalisis pola kepuasan mahasiswa terhadap kinerja pengajaran dosen dengan mengintegrasikan algoritma K-Means Clustering dan algoritma Random Forest. Data penelitian komprehensif yang digunakan mencakup 5.260 observasi individu, yang dianalisis secara sistematis berdasarkan dimensi kualitas layanan dan atribut akademik mahasiswa. Hasil analisis, yang ditentukan dengan memanfaatkan pendekatan Elbow Method, telah berhasil menetapkan tiga kluster optimal. Kluster-kluster tersebut merepresentasikan segmen yang berbeda: segmen Sangat Puas dengan skor 4,84, segmen Cukup Puas dengan skor 4,04, dan segmen Kurang Puas dengan skor 3,04. Selanjutnya, algoritma Random Forest menunjukkan tingkat akurasi prediksi sebesar 47% ketika dievaluasi pada 1.576 sampel data uji. Algoritma tersebut berhasil mengidentifikasi bahwa Beban SKS Semester adalah faktor determinan paling dominan yang memengaruhi kepuasan mahasiswa, dengan nilai kepentingan mencapai 47,46%. Sebuah temuan unik mendemonstrasikan bahwa mahasiswa dengan beban kerja akademik tertinggi, yaitu rata-rata 20,81 SKS, secara aktual diposisikan di dalam segmen Sangat Puas. Pada akhirnya, studi ini menyimpulkan bahwa tingkat keterlibatan akademik mahasiswa yang lebih intensif akan berkorelasi positif dengan apresiasi terhadap kualitas pengajaran dosen. Hasil penelitian ini dapat memberikan panduan strategis bagi manajemen universitas yang bertujuan untuk meningkatkan kualitas layanan berdasarkan pada karakteristik profil akademik mahasiswa

Kata Kunci: Beban SKS, Kepuasan Mahasiswa, Kinerja Dosen, K-Means Clustering, Random Forest.

I. PENDAHULUAN

Kepuasan mahasiswa terhadap dosen merupakan salah satu indikator yang secara luas digunakan untuk menilai kualitas proses pembelajaran di perguruan tinggi[1]. Instrumen evaluasi yang umum digunakan adalah survei dengan skala Likert, namun analisis yang dilakukan sering kali terbatas pada statistik deskriptif sehingga belum mampu menggali pola, struktur, atau segmentasi kepuasan secara lebih mendalam[2]. Padahal, pemahaman yang komprehensif terhadap heterogenitas respon mahasiswa berpotensi memberikan dasar empiris yang lebih kuat untuk peningkatan mutu pengajaran. Metode pelaporan yang hanya mengandalkan nilai rata-rata (*mean*) memiliki

kelemahan karena cenderung menyamakan variasi respons mahasiswa yang sangat besar. Selain itu, belum diketahui secara pasti sejauh mana variabel akademik seperti beban SKS, jumlah mata kuliah, atau tingkat semester memengaruhi terbentuknya segmentasi kepuasan tersebut. Berbagai penelitian terdahulu menunjukkan bahwa tingkat kepuasan mahasiswa dipengaruhi oleh kualitas interaksi pedagogis, kejelasan penyampaian materi, serta integrasi teknologi pendukung pembelajaran [3], [4].

Meskipun demikian, kajian mengenai kepuasan mahasiswa masih menunjukkan adanya kesenjangan penelitian, khususnya terkait minimnya penerapan pendekatan analitik berbasis *machine learning* untuk melakukan segmentasi dan mengidentifikasi faktor determinan kepuasan secara simultan. Sebagian besar penelitian sebelumnya masih berfokus pada pemanfaatan analisis regresi[5] atau uji hubungan sederhana[6]. Pendekatan ini memiliki keterbatasan dalam memetakan profil kelompok mahasiswa yang memiliki karakteristik serupa secara otomatis. Di sisi lain, dalam domain *Educational Data Mining* (EDM), penggunaan algoritma clustering seperti K-Means mulai banyak digunakan untuk segmentasi profil akademik. Namun, penelitian tersebut umumnya berhenti pada tahap pengelompokan tanpa mengevaluasi faktor determinan yang menyebabkan terbentuknya segmen tersebut secara sistematis. Dengan demikian, terdapat kebutuhan untuk mengembangkan pendekatan yang mampu mengidentifikasi kelompok-kelompok kepuasan secara objektif dan memahami faktor akademik yang berkontribusi terhadap pembentukannya.

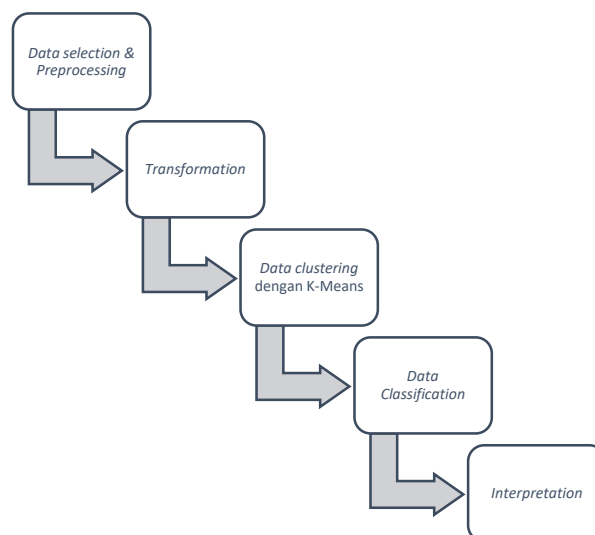
Dalam konteks tersebut, penggunaan metode *machine learning* menjadi relevan dan strategis. Algoritma K-Means sering digunakan untuk pengelompokan profil mahasiswa[7], namun jarang yang melanjutkan analisisnya untuk mencari faktor penyebab di balik terbentuknya kelompok tersebut. Di sisi lain, algoritma Random Forest memiliki keunggulan dalam menentukan tingkat kepentingan fitur (*feature importance*)[8], namun penggunaannya dalam mengevaluasi faktor akademik terhadap tingkat kepuasan mahasiswa masih terbatas[9]. Penelitian ini mencoba mengisi celah tersebut dengan menerapkan pendekatan integratif. Algoritma K-Means memungkinkan identifikasi segmentasi kepuasan berdasarkan kesamaan pola respons, sedangkan Random Forest dapat digunakan untuk menentukan variabel akademik yang memiliki kontribusi terbesar terhadap keanggotaan kluster[10], [11]. Urgensi integrasi pendekatan analitik tersebut semakin menguat dengan adanya tuntutan kebijakan nasional. Standar Nasional Pendidikan Tinggi (SN-Dikti) mengamanatkan bahwa perguruan tinggi harus melaksanakan penjaminan mutu internal secara berkelanjutan, termasuk melalui pengukuran sistematis terhadap persepsi dan kepuasan mahasiswa pada proses pembelajaran. Selain itu, lembaga akreditasi seperti BAN-PT dan LAM mensyaratkan dokumentasi yang menunjukkan bahwa evaluasi pembelajaran telah dianalisis secara mendalam dan diikuti dengan tindak lanjut berbasis bukti. Dengan demikian, penggunaan metode *machine learning* dalam analisis kepuasan mahasiswa tidak hanya menawarkan kontribusi metodologis, tetapi juga mendukung pencapaian standar mutu dan akreditasi pendidikan tinggi di Indonesia [12], [13].

Penelitian ini mengintegrasikan dua tahap analisis utama. Tahap pertama dilakukan dengan mengelompokkan mahasiswa berdasarkan dimensi kualitas layanan (SERVQUAL) menggunakan algoritma K-Means Clustering untuk memetakan pola kepuasan yang ada. Berbeda dengan studi sebelumnya yang hanya memandang variabel akademik sebagai data pelengkap, penelitian ini memposisikan variabel tersebut sebagai faktor prediktif untuk membedah struktur kepuasan mahasiswa. Keunggulan integrasi ini memungkinkan identifikasi kelompok kepuasan secara objektif sekaligus menentukan tingkat kepentingan fitur (*feature importance*) dari variabel seperti beban SKS, jumlah mata kuliah, dan tingkat semester dalam membentuk persepsi mahasiswa. Hasil pemetaan tersebut kemudian digunakan sebagai basis bagi algoritma Random Forest untuk mengidentifikasi variabel akademik yang paling dominan dalam menentukan keanggotaan setiap kluster. Melalui pendekatan ini, analisis kepuasan mahasiswa menjadi lebih mendalam karena tidak hanya berhenti pada pengelompokan, tetapi juga mendukung pengambilan keputusan berbasis data (*data-driven decision-making*) yang lebih presisi [14], [15]. Dengan demikian, penelitian ini bertujuan untuk melakukan segmentasi kepuasan mahasiswa sekaligus memetakan faktor akademik yang paling berpengaruh, sehingga dapat memberikan rekomendasi strategis bagi universitas dalam meningkatkan kualitas layanan instruksional sesuai dengan karakteristik spesifik dari setiap kelompok mahasiswa.

II. METODE

Penelitian ini menggabungkan analisis klusterisasi untuk segmentasi mahasiswa dan analisis klasifikasi untuk identifikasi variabel akademik yang paling berpengaruh terhadap tingkat kepuasan tersebut. Data yang digunakan dalam penelitian ini adalah data isian EDM dari mahasiswa terhadap satu dosen untuk satu mata kuliah spesifik. Dengan demikian, total 5.260 baris data observasi mencerminkan akumulasi persepsi mahasiswa terhadap kinerja instruksional dosen pada periode semester yang berjalan. Atribut data yang digunakan meliputi variabel kepuasan dan variabel akademik. Variabel kepuasan digunakan untuk input klusterisasi sedangkan variabel akademik digunakan untuk input klasifikasi. Variabel Kepuasan (Input Kluster) yang berdasarkan dimensi SERVQUAL yang mencakup *Reliability* (Keandalan), *Responsiveness* (Daya Tanggap), dan *Assurance* (Jaminan). Variabel Akademik (Input Klasifikasi) yaitu: Semester, Beban SKS Semester, dan Jumlah Mata Kuliah yang diambil dalam satu semester[16]. Adapun prosedur dalam penelitian ini dilakukan melalui tahapan yang terstruktur untuk menjamin validitas model yang dihasilkan. Tahapan tersebut digambarkan dalam diagram alir pada gambar 1.

Berikut penjelasan pada gambar 1: Pada tahap pertama dilakukan *data selection & preprocessing*: Data mentah hasil kuesioner dibersihkan dari observasi yang tidak lengkap (*missing values*). Tahap ini krusial untuk memastikan data observasi siap diolah. Mengingat adanya potensi ketidakseimbangan jumlah data antar tingkat kepuasan (*imbalanced data*), dilakukan pengecekan distribusi label hasil klusterisasi. Jika ditemukan ketimpangan signifikan, dilakukan penyeimbangan kelas pada training set untuk memastikan model Random Forest tidak bias terhadap kelas mayoritas. Selanjutnya dilakukan *transformation* yaitu normalisasi data menggunakan *Min-Max Scaling*. Proses ini mengubah seluruh skor dimensi SERVQUAL dan variabel akademik ke dalam rentang 0 hingga 1 agar perbedaan skala antar variabel tidak memengaruhi kinerja algoritma K-Means. Tahapan selanjutnya Adalah proses klusterisasi (*data clustering*) dengan mengimplementasikan algoritma K-Means untuk membagi mahasiswa ke dalam ‘K’ kelompok. Penentuan jumlah kluster terbaik dilakukan dengan menggunakan Elbow Method yaitu dengan mengamati titik siku pada grafik *Within-Cluster Sum of Square* (WCSS). Setelah dilakukan label kluster terbentuk pada tahap *clustering*, data kemudian akan diklasifikasikan pada tahapan *classification*. Klasifikasi data menggunakan label kluster hasil K-Means sebagai variabel target, kemudian menerapkan algoritma Random Forest untuk mengidentifikasi faktor akademik yang paling berpengaruh. Proses klasifikasi menggunakan algoritma Random Forest dilakukan dengan membagi data observasi secara acak menggunakan proporsi 80% sebagai data latih (training set) sebanyak 4.208 sampel dan 20% sebagai data uji (testing set) sebanyak 1.052 sampel. Pembagian ini bertujuan untuk melatih model sekaligus memvalidasi kemampuannya dalam mengklasifikasikan data baru yang belum pernah diolah sebelumnya. Tahap akhir penelitian adalah *Interpretation* yaitu mengvaluasi model dilakukan menggunakan *Confusion Matrix* untuk menghitung nilai *Precision*, *Recall*, dan *F1-Score*. Kontribusi variabel akademik dianalisis melalui fitur *Feature Importance* untuk menentukan variabel mana yang paling dominan dalam membentuk segmentasi kepuasan mahasiswa.



Gambar 1. Alur prosedur penelitian

III. HASIL DAN PEMBAHASAN

A. Data Penelitian

Data yang digunakan pada penelitian ini sebanyak 5.260 data yang terdiri dari persepsi mahasiswa terhadap 13 butir pertanyaan evaluasi yang dibagi menjadi sepuluh pertanyaan kualitas layanan dan tiga pertanyaan tentang dimensi akademik. Pertanyaan kualitas layanan yang berjumlah sepuluh pertanyaan kemudian dikelompokkan ke dalam lima dimensi SERVQUAL, yaitu Reliability, Responsiveness, Assurance, Empathy dan Tangible. Pengelompokan ini bertujuan untuk menyederhanakan input fitur pada algoritma K-Means tanpa menghilangkan esensi dari kualitas layanan dosen. Pemetaan pertanyaan penelitian ke dalam dimensi SERVQUAL ditunjukkan pada tabel 1. Kategori kualitas layanan untuk input K-Means dibagi menjadi lima dimensi yang terdiri dari beberapa pertanyaan untuk setiap dimensinya. Dimensi Reliability merupakan kemampuan dosen untuk memberikan layanan yang dijanjikan secara akurat, tepat waktu, dan memuaskan. Hal ini merupakan wujud dari profesionalisme seorang dosen seperti ketepatan waktu mengajar dan kesesuaian materi dengan Rencana Pembelajaran Semester (RPS). Pertanyaan yang relevan dengan dimensi ini yaitu terkait X1: Keandalan dosen dalam menyampaikan materi sesuai RPS dan X2: Kedisiplinan dan ketepatan waktu kehadiran dosen. Dimensi Responsiveness merupakan kemauan untuk membantu pelanggan (mahasiswa) dan memberikan layanan yang cepat atau sigap. Dimensi ini fokus pada bagaimana dosen merespons pertanyaan, kesulitan mahasiswa, atau memberikan umpan balik atas tugas. Pertanyaan yang relevan

dengan dimensi ini yaitu terkait X3: Kecepatan dosen dalam merespons pertanyaan mahasiswa dan X4: Kesiadaan dosen membantu kesulitan akademik mahasiswa. Dimensi Assurance terkait pengetahuan, kesopanan, dan kemampuan personel untuk menumbuhkan rasa percaya dan keyakinan. Hal ini terkait dengan kompetensi akademik dosen (penguasaan materi) dan cara dosen menciptakan rasa aman/nyaman saat berdiskusi. Pertanyaan yang relevan dengan dimensi ini yaitu terkait X5: Penguasaan materi dan kompetensi keilmuan dosen dan X6: Kemampuan dosen menciptakan rasa aman dan percaya diri. Dimensi Empathy terkait kepedulian dan perhatian dosen yang diberikan kepada mahasiswa seperti bagaimana dosen memahami kondisi mahasiswa, bersikap adil, dan memberikan perhatian secara personal tanpa diskriminasi. Pertanyaan yang relevan dengan dimensi ini yaitu X7: Kepedulian dosen terhadap kondisi/kebutuhan mahasiswa dan X8: Kemudahan komunikasi dan kedekatan hubungan dosen-mahasiswa. Dimensi Tangibles berfokus pada tampilan fisik, fasilitas, peralatan, dan penampilan dosen yang mendukung layanan. Dalam proses kegiatan belajar mengajar, ini berkaitan dengan media pembelajaran, kebersihan kelas, kelengkapan alat bantu (proyektor/lab), serta kerapian dosen. Pertanyaan yang relevan dengan dimensi ini yaitu X9: Kerapian penampilan dosen saat mengajar dan X10: Pemanfaatan media/alat bantu pembelajaran yang memadai.

TABEL 1
PEMETAAN PERTANYAAN PENELITIAN KE DALAM DIMENSI SERVQUAL

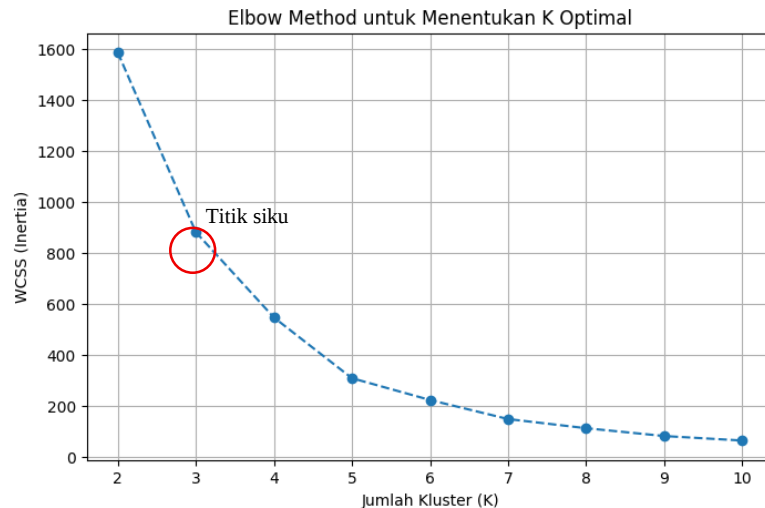
Kategori	Dimensi	Kode	Indikator / Butir Pertanyaan Kuesioner	Skala
Kualitas Layanan (Input K-Means)	Reliability	X1	Keandalan dosen dalam menyampaikan materi sesuai RPS.	Likert 1-5
		X2	Kedisiplinan dan ketepatan waktu kehadiran dosen.	Likert 1-5
	Responsiveness	X3	Kecepatan dosen dalam merespons pertanyaan mahasiswa.	Likert 1-5
		X4	Kesiadaan dosen membantu kesulitan akademik mahasiswa.	Likert 1-5
	Assurance	X5	Penguasaan materi dan kompetensi keilmuan dosen.	Likert 1-5
		X6	Kemampuan dosen menciptakan rasa aman dan percaya diri.	Likert 1-5
	Empathy	X7	Kepedulian dosen terhadap kondisi/kebutuhan mahasiswa.	Likert 1-5
		X8	Kemudahan komunikasi dan kedekatan hubungan dosen-mahasiswa.	Likert 1-5
	Tangibles	X9	Kerapian penampilan dosen saat mengajar.	Likert 1-5
		X10	Pemanfaatan media/alat bantu pembelajaran yang memadai.	Likert 1-5
Data Akademik (Input Random Forest)	Akademik	A1	Tingkat Semester (Semester).	Rasio
		A2	Beban SKS Semester (SKS).	Rasio
		A3	Jumlah Mata Kuliah (MK).	Rasio

Berdasarkan Tabel 1, variabel X1 hingga X10 merupakan data persepsi yang digunakan sebagai fitur masukan (*input feature*) untuk algoritma K-Means Clustering guna menentukan segmen kepuasan. Sementara itu, variabel A1 hingga A3 merupakan atribut akademik mahasiswa yang digunakan sebagai variabel prediktor dalam model Random Forest untuk mengidentifikasi faktor yang paling memengaruhi klasifikasi kepuasan tersebut.

B. Clustering dengan K-Means

1. Penentuan Jumlah Kluster

Sebelum dilakukan proses klusterisasi, dilakukan penentuan jumlah kluster optimal (K) menggunakan *Elbow Method*[17]. Berdasarkan gambar 2, terlihat bahwa penurunan nilai *Within-Cluster Sum of Squares* (WCSS) mulai melandai secara signifikan pada titik K =3. Titik 'siku' (*elbow*) ini menunjukkan bahwa penambahan jumlah kluster di atas tiga tidak lagi memberikan penurunan variansi yang signifikan dalam kelompok. Oleh karena itu, jumlah kluster yang ditetapkan dalam penelitian ini adalah tiga segmen kepuasan.



Gambar 2. Elbow Method

2. Segmen Karakteristik Kepuasan

Berdasarkan hasil pemetaan centroid pada Tabel 2, terbentuk tiga segmen kepuasan mahasiswa dengan karakteristik kuantitatif yang berbeda secara signifikan.

TABEL 2
KARAKTERISTIK CENTROID DIMENSI SERVQUAL PER KLASSTER

Dimensi	Klaster 1 (Sangat Puas)	Klaster 2 (Cukup Puas)	Klaster 3 (Kurang Puas)
Tangibles	4.82	4.02	3.01
Reliability	4.86	4.06	3.06
Responsiveness	4.85	4.05	3.05
Assurance	4.84	4.04	3.04
Empathy	4.83	4.03	3.03
Rata-rata Skor	4.84	4.04	3.04

Klaster pertama diidentifikasi sebagai Segmen Kepuasan Tinggi, yang mencakup responden dengan persepsi positif yang konsisten di seluruh dimensi layanan dengan rata-rata skor 4,84. Karakteristik utama segmen ini adalah tingginya nilai pada dimensi *Reliability* (4,86) dan *Responsiveness* (4,85), yang menunjukkan bahwa mahasiswa dalam kelompok ini merasakan tingkat profesionalisme, ketepatan waktu, dan daya tanggap dosen yang sangat tinggi dalam memberikan umpan balik. Mahasiswa dalam kelompok ini merepresentasikan pengguna yang sangat puas dan dapat dijadikan sebagai *benchmark* keberhasilan layanan pengajaran dosen yang ideal bagi program studi.

Sebaliknya, Segmen Kepuasan Sedang menjadi kelompok mayoritas yang mencerminkan pengalaman layanan cukup memadai namun belum optimal dengan rata-rata skor 4,04. Meskipun secara umum mahasiswa dalam kelompok ini sudah merasa puas, terdapat selisih skor (*gap*) sekitar 0,80 poin jika dibandingkan dengan klaster pertama. Skor terendah pada kelompok ini ditemukan pada dimensi *Tangibles* (4,02), yang mengindikasikan bahwa ketersediaan media pembelajaran dan sarana pendukung kelas menjadi aspek utama yang perlu ditingkatkan untuk mendorong status kepuasan mereka ke tingkat yang lebih tinggi.

Sementara itu, Segmen Kepuasan Rendah merupakan kelompok minoritas namun memerlukan perhatian prioritas. Klaster ini ditandai dengan skor rendah secara sistematis pada seluruh dimensi, terutama pada aspek *Tangibles* (3,01) dan *Empathy* (3,03). Rendahnya skor pada dimensi empati menunjukkan adanya hambatan fundamental dalam interaksi personal antara dosen dan mahasiswa pada kelompok ini. Hal ini menjadi sinyal kritis bagi manajemen universitas bahwa perbaikan tidak hanya diperlukan pada aspek fasilitas fisik, tetapi juga pada pendekatan pedagogis yang lebih humanis dan adil guna mencegah penurunan motivasi belajar mahasiswa pada segmen tersebut.

3. Analisis Kinerja Model dan Identifikasi Faktor Kepuasan

Tahap ini bertujuan untuk memverifikasi kemampuan model dalam mengenali pola kepuasan serta menentukan variabel yang paling berpengaruh. Pengujian dilakukan menggunakan algoritma Random Forest dengan membagi data menjadi *training set* dan *testing set*[18].

Performa model dalam memprediksi keanggotaan klaster dievaluasi menggunakan metrik *precision*, *recall*, dan *f1-score*. Berdasarkan hasil pengujian pada 1.576 data uji, model menghasilkan akurasi keseluruhan sebesar 47% yang ditunjukkan pada tabel 3.

TABEL 3
HASIL ANALISIS RANDOM FOREST

Klaster Kepuasan	Precision	Recall	F1-Score	Support
0 (Sedang)	0.66	0.40	0.50	888
1 (Tinggi)	0.47	0.58	0.52	628
2 (Rendah)	0.06	0.25	0.09	60
Accuracy			0.47	1576
Macro Avg	0.40	0.41	0.37	1576
Klaster Kepuasan	Precision	Recall	F1-Score	Support
Weighted Avg	0.56	0.47	0.49	1576

Tabel 3 menunjukkan bahwa model memiliki kemampuan prediksi yang cukup stabil pada Klaster 0 (Sedang) dengan precision 0,66. Rendahnya performa pada Klaster 2 disebabkan oleh ketidakseimbangan data (*imbalanced data*), di mana jumlah sampel mahasiswa tidak puas sangat sedikit (60 observasi). Namun secara *weighted average*, model masih mampu memberikan gambaran pola yang representatif bagi mayoritas data.

Meskipun akurasi model berada pada tingkat moderat, algoritma Random Forest berhasil mengekstraksi tingkat kepentingan fitur yang memberikan informasi mengenai faktor utama pembentuk kepuasan mahasiswa.

TABEL 4
PERINGKAT KEPENTINGAN FITUR

Peringkat	Faktor Akademik	Importance (%)	Importance_pct
2	Beban SKS Semester	0.47	47,46%
1	Jumlah Mata Kuliah	0.38	37,91%
0	Semester	0.15	14,63%

Hasil pada Tabel 4 menunjukkan bahwa Beban SKS merupakan variabel yang paling dominan dalam membedakan antar klaster dengan kontribusi sebesar 47,46%. Hal ini menegaskan bahwa beban studi yang diambil mahasiswa secara signifikan lebih berpengaruh terhadap persepsi mereka pada kinerja dosen dibandingkan faktor masa studi atau jumlah mata kuliah.

Performa model Random Forest menghasilkan akurasi keseluruhan sebesar 47%, yang menunjukkan tantangan signifikan dalam memprediksi keanggotaan klaster secara presisi berdasarkan atribut akademik saja. Meskipun pada bagian metodologi telah direncanakan langkah pengecekan dan penyeimbangan kelas (*class balancing*) untuk memitigasi potensi imbalanced data, hasil pengujian pada Tabel 3 mengonfirmasi bahwa model tetap memiliki keterbatasan dalam mengenali Klaster 3 (Tidak Puas) dengan nilai F1-score yang sangat rendah (0,09). Hal ini mengindikasikan bahwa ketimpangan jumlah data dalam populasi isian EDOM sangat ekstrem, di mana mayoritas responden cenderung memberikan penilaian pada kelas puas, sehingga model tetap memiliki kecenderungan bias terhadap kelas mayoritas meskipun upaya penyeimbangan telah dilakukan. Rendahnya performa pada segmen minoritas ini juga memperkuat dugaan bahwa terdapat variabel laten lain di luar atribut akademik yang lebih dominan dalam membentuk ketidakpuasan mahasiswa, sehingga hasil feature importance yang menempatkan SKS sebagai variabel utama (47,46%) sebaiknya diinterpretasikan sebagai temuan eksploratif yang memerlukan validasi lebih lanjut menggunakan teknik penanganan data yang lebih kompleks seperti SMOTE atau algoritma klasifikasi yang lebih robust terhadap data yang tidak seimbang.

4. Profiling karakteristik mahasiswa

TABEL 5
TABEL HASIL PROFILING SEGMENT

Profil Segmen	Rata-rata Kepuasan	Skor	Beban SKS Semester	Jumlah MK Diambil	Semester
Segmen Sangat Puas	4.84	21	10	3	
Segmen Cukup Puas	4.04	20	9	3	
Segmen Tidak Puas	3.04	19	9	3	

Berdasarkan Tabel 5, dapat terlihat profiling mahasiswa pada setiap segmen yang memberikan gambaran lebih mendalam mengenai karakteristik akademik di balik setiap klaster.

4.1 Segmen Sangat Puas (Skor 4,84)

Kelompok ini merupakan mahasiswa dengan beban akademik tertinggi, yaitu rata-rata 21 SKS dan 10 mata kuliah. Temuan ini menunjukkan bahwa mahasiswa yang memiliki keterlibatan akademik tinggi justru cenderung memberikan penilaian yang sangat positif terhadap kualitas pengajaran dosen. Secara teknis, dominansi jumlah sampel pada segmen ini juga menjelaskan mengapa model Random Forest memiliki *recall* yang lebih baik pada kelas ini, karena model lebih mudah mengenali pola dari kelompok mayoritas yang memiliki keterlibatan akademik intensif.

4.2 Segmen Cukup Puas (Skor 4,04)

Kelompok ini berada pada posisi moderat dengan rata-rata 20 SKS. Meskipun beban SKS-nya mirip dengan segmen sangat puas, jumlah mata kuliah yang diambil lebih sedikit (9 SKS), yang menunjukkan adanya perbedaan fokus atau struktur kurikulum pada kelompok ini.

4.3 Segmen Kurang Puas (Skor 3,04)

Mahasiswa pada segmen ini memiliki beban akademik terendah dengan rata-rata 19 SKS. Hal ini mengindikasikan bahwa ketidakpuasan lebih sering muncul pada mahasiswa dengan beban studi yang relatif lebih ringan, yang mungkin disebabkan oleh kurangnya intensitas interaksi dengan layanan akademik yang tersedia. Namun, sebagaimana telah disinggung pada evaluasi performa model, interpretasi terhadap segmen ini harus dilakukan secara hati-hati sebagai temuan eksploratif. Hal ini dikarenakan adanya kendala *imbalanced data* di mana jumlah sampel pada segmen ini sangat sedikit dalam populasi yang menyebabkan model klasifikasi memiliki keterbatasan dalam memetakan variabel determinan secara presisi untuk kelompok minoritas ini.

IV. SIMPULAN

Penelitian ini mampu memetakan struktur kepuasan mahasiswa terhadap kinerja pengajaran dosen dengan mengintegrasikan algoritma K-Means Clustering dan Random Forest. Berdasarkan analisis terhadap 5.260 data observasi, ditemukan tiga segmen kepuasan yang optimal ($K=3$), yaitu: Sangat Puas (skor 4,84), Cukup Puas (skor 4,04), dan Kurang Puas (skor 3,04). Pada tahap klasifikasi, model Random Forest menghasilkan akurasi keseluruhan sebesar 47%. Meskipun performa prediksi pada segmen minoritas (Kurang Puas) masih rendah akibat kendala *imbalanced data*, model ini memberikan temuan eksploratif yang signifikan mengenai kontribusi atribut akademik. Variabel Beban SKS Semester teridentifikasi sebagai faktor yang paling berpengaruh dalam membedakan segmen kepuasan dengan tingkat kepentingan sebesar 47,46%, disusul oleh Jumlah Mata Kuliah (37,91%). Penelitian ini mengungkap fenomena unik di mana mahasiswa dengan beban akademik tertinggi (rata-rata 21 SKS) secara konsisten berada pada segmen "Sangat Puas". Temuan ini mengindikasikan bahwa intensitas keterlibatan akademik berkorelasi positif dengan apresiasi mahasiswa terhadap kualitas instruksional dosen. Sebaliknya, ketidakpuasan cenderung ditemukan pada mahasiswa dengan beban studi yang lebih ringan (19 SKS), yang diduga berkaitan dengan minimnya interaksi akademik. Oleh karena itu, manajemen universitas disarankan untuk mengoptimalkan strategi keterlibatan mahasiswa melalui pengelolaan beban akademik dan distribusi mata kuliah yang seimbang. Terlepas dari kontribusi yang dihasilkan, rendahnya performa model pada segmen minoritas menjadi catatan kritis dalam penelitian ini. Untuk penelitian selanjutnya, disarankan menerapkan teknik penanganan *imbalanced data* yang lebih lanjut, seperti *Synthetic Minority Over-sampling Technique* (SMOTE), serta mengeksplorasi penggunaan algoritma yang lebih robust seperti XGBoost. Selain itu, pengembangan model dengan memasukkan variabel non-akademik, seperti latar belakang sosial-ekonomi dan kualitas fasilitas digital, diharapkan dapat meningkatkan akurasi klasifikasi serta memberikan pemahaman yang lebih holistik mengenai determinan kepuasan mahasiswa.

UCAPAN TERIMAKASIH

Terima kasih kepada Lembaga Penelitian dan Pengabdian Masyarakat (LPPM) Universitas Tidar yang telah memberikan hibah Dana DIPA Penelitian Afirmasi Asisten Ahli tahun 2025.

DAFTAR PUSTAKA

- [1] A. Rusdi Widya and C. Author, "Tingkat Kepuasan Mahasiswa dalam Pembelajaran di Program Studi Teknik Industri: Studi Kasus Perguruan Tinggi Swasta di Jawa Barat," *R2J*, vol. 7, no. 3, 2025, doi: 10.38035/rj.v7i3.
- [2] E. P. Setiawan, "STATISTICAL LITERACY IN PRIMARY SCHOOL MATHEMATICS CURRICULA: HISTORICAL REVIEW AND DEVELOPMENT," *Jurnal Pendidikan dan Kebudayaan*, vol. 6, no. 1, pp. 1–20, Jun. 2021, doi: 10.24832/jpnk.v6i1.1915.
- [3] L. Sari, A. Romadloni, and R. Listyaningrum, "Penerapan Data Mining dalam Analisis Prediksi Kanker Paru Menggunakan Algoritma Random Forest," *Infotekmesin*, vol. 14, no. 1, pp. 155–162, Jan. 2023, doi: 10.35970/infotekmesin.v14i1.1751.
- [4] U. Khultsum and A. Subekti, "Penerapan Algoritma Random Forest dengan Kombinasi Ekstraksi Fitur Untuk Klasifikasi Penyakit Daun Tomat," *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 5, no. 1, p. 186, Jan. 2021, doi: 10.30865/mib.v5i1.2624.
- [5] D. D. Saputra, "ANALISIS PERBAIKAN KUALITAS PELAYANAN UNTUK MENINGKATKAN KEPUASAN PELANGGAN DENGAN MENGGUNAKAN METODE SERVQUAL, IPA, DAN REGRESI LINEAR BERGANDA," *Inaque : Journal of Industrial and Quality Engineering*, 2022, [Online]. Available: <https://api.semanticscholar.org/CorpusID:250394756>
- [6] Y. Rahesa Wijaya and E. Tjahjaningsih, "Analisis Faktor Kepuasan Pelanggan Serta Dampaknya Pada Niat Beli Ulang," 2022. [Online]. Available: <http://journal.stekom.ac.id/index.php/E-Bisnis/page201>
- [7] Muhammad Zulfadhillah, Mambang, and Septyan Eka Prastya, "Implementasi Metode K-Means Clustering untuk Meningkatkan Penjangkauan Mahasiswa," *TEMATIK*, vol. 9, no. 2, pp. 152–160, Nov. 2022, doi: 10.38204/tematik.v9i2.1053.
- [8] M. A. Pandu W, R. E. Saputro, P. Purwadi, and U. A. Rohmah, "Analisis Tingkat Akurasi Metode Naive Bayes dan Random Forest dalam Prediksi Penjualan Emas," *Jurnal Pendidikan dan Teknologi Indonesia*, vol. 5, no. 7, pp. 1809–1821, Jul. 2025, doi: 10.52436/1.jpti.732.
- [9] S. Sobari, A. I. Purnamasari, A. Bahtiar, and K. Kaslani, "MENINGKATKAN MODEL PREDIKSI KELULUSAN SANTRI TAHFIDZ DI PONDOK PESANTREN AL-KAUTSAR MENGGUNAKAN ALGORITMA RANDOM FOREST," *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 13, no. 1, Jan. 2025, doi: 10.23960/jitet.v13i1.5704.
- [10] D. Eko Waluyo *et al.*, "Implementasi Algoritma Regresi pada Machine Learning untuk Prediksi Indeks Harga Saham Gabungan," *Universitas Dian Nuswantoro, Semarang Jln. Imam Bonjol*, vol. 9, no. 1.
- [11] H. Argretya and K. Kusnawi, "Impelementasi Random Oversampling pada Algoritma Random Forest untuk Klasifikasi Indeks Standar Udara di DKI Jakarta 2024," *Jurnal Informatika: Jurnal Pengembangan IT*, vol. 10, no. 4, pp. 1189–1198, Sep. 2025, doi: 10.30591/jpit.v10i4.9466.
- [12] T. Hartati, O. Nurdian, E. Wiyandi, and S. I. Cirebon, "ANALISIS DAN PENERAPAN ALGORITMA K-MEANS DALAM STRATEGI PROMOSI KAMPUS AKADEMI MARITIM SUAKA BAHARI," 2021.
- [13] C. Kamila, M. Adiyatma, G. R. Namang, R. Ramadhan, F. Syah, and D. Redaksi, "Pendidikan Teknik Informatika dan Komputer, Fakultas Teknik," *JURNAL INTECH*, vol. 2, no. 1, p. 1.
- [14] O. Pahlevi and Y. Handrianto, "Implementasi Algoritma Klasifikasi Random Forest Untuk Penilaian Kelayakan Kredit." [Online]. Available: <http://ejournal.bsi.ac.id/ejurnal/index.php/infotech>
- [15] I. Afdhal *et al.*, "Penerapan Algoritma Random Forest Untuk Analisis Sentimen Komentar Di YouTube Tentang Islamofobia," *Jurnal Nasional Komputasi dan Teknologi Informasi*, vol. 5, no. 1, 2022.
- [16] S. Daryazadeh, M. Yavari, M. R. Sharif, M. J. Azadchahr, S.-V. Zabihi Hoseini, and H. Akbari, "Assessing the Quality of Educational Services from the Viewpoint of Clinical Teachers and Medical Students Using SERVQUAL Model," *Educational Research in Medical Sciences*, vol. 11, no. 2, Jan. 2023, doi: 10.5812/erms-119417.
- [17] N. A. Maori, "METODE ELBOW DALAM OPTIMASI JUMLAH CLUSTER PADA K-MEANS CLUSTERING," *Jurnal SIMETRIS*, vol. 14, 2023.
- [18] I. Septiana and D. Alita, "Perbandingan Random Forest dan SVM dalam Analisis Sentimen Quick Count Pemilu 2024," *Jurnal Informatika: Jurnal Pengembangan IT*, vol. 9, no. 3, pp. 224–233, Dec. 2024, doi: 10.30591/jpit.v9i3.6640.