

# Pemanfaatan Algoritma *Random Forest Regression* dalam Memprediksi Kepuasan Mahasiswa Terhadap Dosen

Rostika Listyaningrum<sup>1\*</sup>, Riyadi Purwanto<sup>2</sup>, Dwi Novia Prasetyanti<sup>3</sup>,  
Cahya Vikasari<sup>4</sup>, Ardhita Fajar Pratiwi<sup>5</sup>

<sup>1, 2, 3, 4</sup>Program Penelitian Teknik Informatika, Politeknik Negeri Cilacap

<sup>5</sup>Program Penelitian Teknologi Rekayasa Mekatronika, Politeknik Negeri Cilacap

<sup>1, 2, 3, 4, 5</sup>Jln. Dr. Soetomo No.1 Karangcengis Sidakaya, Kabupaten Cilacap, 53212, Indonesia

E-mail: li\_sa007@pnc.ac.id<sup>1</sup>, riyadi\_purwanto@pnc.ac.id<sup>2</sup>, dnprasetyanti@pnc.ac.id<sup>3</sup>, cahyavikasari@pnc.ac.id<sup>4</sup>, ardhita@pnc.ac.id<sup>5</sup>

---

## Abstrak

---

### Info Naskah:

Naskah masuk: 9 Juni 2025

Direvisi: 7 Juli 2025

Diterima: 14 Juli 2025

Kepuasan mahasiswa terhadap dosen merupakan indikator utama dalam menilai kualitas pembelajaran di perguruan tinggi. Sayangnya, pendekatan evaluasi yang umum digunakan masih bersifat deskriptif dan subjektif, sehingga kurang efektif untuk mendukung perbaikan berkelanjutan. Selain itu, pemanfaatan indikator kompetensi dosen secara menyeluruh dalam model prediksi masih terbatas. Penelitian ini mengisi kekosongan tersebut dengan mengembangkan model prediksi kepuasan mahasiswa menggunakan algoritma *Random Forest Regression* yang dioptimalkan melalui *grid search* dan seleksi fitur dengan metode *Recursive Feature Elimination* yang dikombinasikan dengan validasi silang *5-fold*. Data diperoleh dari sistem EDOM Politeknik Negeri Cilacap, mencakup 24 indikator berdasarkan standar kompetensi dosen dan dianalisis menggunakan perangkat lunak R. Model terbaik diperoleh dengan parameter  $mtry = 1$  dan  $ntree = 300$ , menghasilkan nilai RMSE 0,0222, MAE 0,0118, dan  $R^2$  0,9959. Tiga indikator utama yang paling berpengaruh adalah pemberian tugas terstruktur, keragaman metode pembelajaran, dan ketepatan kehadiran. Hasil ini diharapkan menjadi dasar kebijakan peningkatan mutu pendidikan tinggi secara berkelanjutan.

---

## Abstract

---

### Keywords:

student satisfaction;  
lecturer performance;  
random forest regression;  
recursive feature elimination;  
higher education quality.

Student satisfaction with lecturers is a key indicator in assessing the quality of higher education. However, commonly used evaluation approaches remain largely descriptive and subjective, making them less effective in supporting sustainable quality improvement. Moreover, the comprehensive use of lecturer competency indicators in predictive models is still limited. This study addresses the gap by developing a student satisfaction prediction model using the Random Forest Regression algorithm, optimized through grid search and feature selection using the Recursive Feature Elimination (RFE) method combined with 5-fold cross-validation. Data were collected from the EDOM system of Politeknik Negeri Cilacap, involving 24 indicators based on national lecturer competency standards, and analyzed using R software. The best model was achieved with parameters  $mtry = 1$  and  $ntree = 300$ , yielding  $RMSE = 0.0222$ ,  $MAE = 0.0118$ , and  $R^2 = 0.9959$ . The three most influential indicators identified were structured assignments, diversity of teaching methods, and punctuality. These findings are expected to inform policies for improving the quality of higher education.

---

\*Penulis korespondensi:

Rostika Listyaningrum

E-mail: li\_sa007@pnc.ac.id

## 1. Pendahuluan

Mutu pendidikan tinggi tidak bisa dilepaskan dari peran sentral dosen dalam proses pembelajaran. Dosen yang baik tidak hanya menyampaikan materi ajar, tetapi juga membentuk pengalaman belajar yang bermakna bagi mahasiswa. Oleh karena itu, evaluasi terhadap dosen menjadi elemen krusial dalam menjaga kualitas akademik dan mendorong profesionalisme di lingkungan perguruan tinggi. Hal ini sejalan dengan amanat Permendikbudristek No. 44 Tahun 2024, yang menekankan bahwa dosen harus memiliki kecakapan tidak hanya secara akademis, tetapi juga dalam menjalin interaksi pembelajaran yang bersifat empatik dan profesional. Setiap pendidik diwajibkan memiliki empat kompetensi utama yang meliputi kemampuan pedagogik, integritas kepribadian, keahlian profesional, serta keterampilan sosial dalam menjalankan tugasnya secara efektif dan bertanggung jawab [1].

Di era digital ini, pendekatan berbasis kecerdasan buatan seperti *machine learning* mulai dilirik sebagai alat bantu untuk mengevaluasi kinerja dosen secara lebih objektif. Salah satu algoritma yang banyak digunakan adalah Random Forest karena kemampuannya dalam mengolah data yang kompleks serta memberikan hasil prediksi yang akurat. Sejumlah penelitian telah menguji efektivitas algoritma ini dalam berbagai konteks pendidikan. Almasri et al. [2] berhasil memprediksi kepuasan mahasiswa dalam pembelajaran daring dengan akurasi hingga 92% menggunakan Random Forest. Widayati et al. [3] mengaplikasikannya untuk mengukur loyalitas mahasiswa, sementara Nachouki et al. [4] menerapkannya dalam prediksi performa akademik mahasiswa. Riyanto et al. [5] bahkan menekankan kemampuan algoritma ini dalam mengekstraksi variabel-variabel penting yang relevan terhadap kepuasan mahasiswa. Selain itu, Sufina dan Wati [6] mulai mengeksplorasi Random Forest untuk mengevaluasi kinerja dosen berbasis Sasaran Kinerja Pegawai (SKP), namun penelitian tersebut masih terbatas pada aspek klasifikasi terhadap dua dimensi kinerja (utama dan perilaku), dan belum menyoroti prediksi kepuasan mahasiswa secara menyeluruh.

Meski begitu, kajian-kajian terdahulu masih menyisakan sejumlah keterbatasan. Umumnya, model yang dibangun berfokus pada klasifikasi (misalnya, puas atau tidak puas), padahal kenyataannya tingkat kepuasan mahasiswa bersifat kontinu dan lebih tepat dianalisis melalui pendekatan regresi. Selain itu, banyak penelitian hanya menggunakan variabel umum seperti IPK, jenis kelamin, atau latar belakang pendidikan [3], [4], [5], tanpa menyoroti secara mendalam indikator kompetensi dosen yang disesuaikan dengan standar nasional. Beberapa penelitian seperti Tiftazani et al. [7] hanya menyoroti aspek layanan akademik, belum menyentuh dimensi kompetensi dosen secara menyeluruh. Rani dan Rakhmawati [8] mengulas kepuasan mahasiswa terhadap kinerja dosen selama perkuliahan daring di masa pandemi, namun pendekatannya masih deskriptif dan belum memanfaatkan model prediktif. Penelitian kuantitatif oleh Selviani et al. [9] menunjukkan bahwa kompetensi pedagogik berpengaruh signifikan terhadap kinerja akademik dosen, tetapi belum diintegrasikan ke dalam sistem prediksi berbasis *machine*

*learning*. Suryanto [10] juga meneliti kepuasan mahasiswa terhadap pengajaran dosen, layanan akademik, dan fasilitas pembelajaran, namun belum secara spesifik mengaitkan indikator kompetensi dosen dengan pendekatan algoritmik atau analitik prediktif.

Pendekatan klasifikasi seperti *Naive Bayes* dan *Decision Tree C4.5* masih mendominasi dalam evaluasi kepuasan mahasiswa, sebagaimana terlihat pada penelitian Ramadani et al. [11] yang memanfaatkan *Naive Bayes* dan menghasilkan akurasi prediksi hingga 98,28% untuk kategori "sangat puas", meskipun penggunaan data terbatas pada parameter umum dan belum melibatkan indikator kompetensi dosen secara utuh. Demikian pula, Fikrul Ilmi et al. [12] menggunakan algoritma C4.5 dan menunjukkan akurasi sebesar 80%, namun penelitian ini hanya mencakup empat variabel (komunikasi, suasana belajar, penilaian, dan penyampaian materi), tanpa memperhatikan struktur kompetensi dosen yang sesuai dengan standar nasional.

Hal lain yang masih kurang mendapatkan perhatian adalah proses tuning parameter pada *Random Forest Regression*. Nilai-nilai penting seperti *mtry* dan *ntree* sering kali dibiarkan pada pengaturan default, tanpa upaya optimalisasi. Padahal, pemilihan parameter yang tepat dapat berpengaruh signifikan terhadap akurasi model prediksi, termasuk penurunan nilai RMSE dan peningkatan  $R^2$  [13]. Penelitian-penelitian sebelumnya menunjukkan bahwa tanpa proses optimasi seperti *grid search* atau validasi silang, potensi algoritma Random Forest tidak dimanfaatkan sepenuhnya [14]. Lebih jauh, hasil penelusuran literatur terbaru menunjukkan bahwa belum ditemukan jurnal yang secara spesifik membahas analisis kepuasan mahasiswa terhadap dosen menggunakan metode *Random Forest Regression*, terutama yang memanfaatkan indikator kompetensi dosen secara menyeluruh sebagaimana diatur dalam kebijakan nasional.

Melihat celah tersebut, penelitian ini mengusulkan pembangunan model prediksi kepuasan mahasiswa terhadap dosen dengan menggunakan *Random Forest Regression* yang ditingkatkan melalui tuning parameter secara sistematis dan seleksi fitur menggunakan metode *Recursive Feature Elimination* (RFE). Data yang digunakan bersumber dari sistem EDOM Politeknik Negeri Cilacap, dengan 24 indikator yang mencakup dimensi kompetensi pedagogik, profesional, sosial, kepribadian, serta aspek sarana dan prasarana. Penelitian ini tidak hanya bertujuan menghasilkan prediksi tingkat kepuasan mahasiswa yang lebih akurat, tetapi juga memberikan wawasan mengenai indikator-indikator mana yang paling berpengaruh terhadap persepsi mahasiswa. Dengan demikian, hasil kajian ini diharapkan dapat menjadi acuan dalam penyusunan kebijakan peningkatan kualitas pembelajaran secara berkelanjutan.

## 2. Metode

Penelitian ini dilaksanakan dengan pendekatan kuantitatif menggunakan metode data mining, memanfaatkan algoritma *Random Forest Regression* untuk memprediksi tingkat kepuasan mahasiswa terhadap dosen. Proses penelitian dirancang secara sistematis ke dalam empat tahapan utama: *data preparation* (persiapan data), *modeling* (pemodelan), *evaluation* (evaluasi model), dan analisis

*feature importance* (Evaluasi Variabel Penting) sebagaimana ditunjukkan pada Gambar 1.



Gambar 1. Tahapan Penelitian

### 2.1 Persiapan Data

Tahap awal dalam proses ini dimulai dengan pengumpulan data, eksplorasi awal, hingga analisis deskriptif untuk memahami karakteristik masing-masing variabel yang akan digunakan dalam model.

#### 2.1.1 Pengumpulan Data

Data yang digunakan dalam penelitian ini bersifat primer, diambil langsung dari sistem Evaluasi Dosen Oleh Mahasiswa (EDOM) pada Semester Ganjil Tahun Akademik 2023/2024 di Politeknik Negeri Cilacap. Total terdapat 1.160 responden mahasiswa yang memberikan penilaian terhadap 70 dosen dari 10 program studi.

Instrumen yang digunakan berupa kuesioner tertutup dengan skala Likert 5 poin, yang dirancang untuk mengukur persepsi mahasiswa terhadap berbagai aspek kinerja dosen. Kuesioner terdiri atas 24 butir pertanyaan (diberi kode I1–I24) yang digunakan sebagai variabel prediktor (X), serta satu variabel target (Y) yang merupakan rata-rata keseluruhan skor kepuasan mahasiswa terhadap dosen. Rentang skor skala Likert adalah sebagai berikut: 1 = Sangat Buruk, 2 = Buruk, 3 = Cukup Baik, 4 = Baik, 5 = Sangat Baik. Data variabel prediktor (X) dan variabel target (Y) ditampilkan dalam tabel berikut:

Tabel 1. Variabel Prediktor (X) dan Variabel Target (Y)

DOSEN	I1	I2	I3	...	I22	I23	I24	Y
1	4,19	4,20	4,14	...	3,97	4,22	4,11	4,11
2	4,17	4,03	4,09	...	3,93	4,05	4,01	4,03
3	4,31	4,38	4,38	...	4,15	4,12	4,19	4,21
4	4,35	4,35	4,35	...	4,07	4,21	4,27	4,24
5	4,18	4,18	4,14	...	3,96	4,16	4,12	4,10
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
65	3,82	3,75	3,61	...	3,60	3,79	3,74	3,72
66	4,16	3,91	3,83	...	3,81	3,85	3,94	3,88
67	3,72	3,89	3,81	...	3,67	3,71	3,68	3,74
68	4,00	4,03	3,77	...	3,77	3,89	3,92	3,83
69	4,19	4,10	4,03	...	4,00	3,96	4,03	4,04
70	4,10	4,12	4,10	...	3,96	3,85	4,07	4,01

Masing-masing butir pertanyaan tersebut dirancang untuk merepresentasikan berbagai kompetensi inti dosen yang telah diatur dalam standar nasional pendidikan tinggi, yaitu:

- Kompetensi Sosial (*empathy*): I1, I3, I4, I5
- Kompetensi Pedagogik (*assurance*): I2, I8, I9, I10, I11, I12, I13, I14, I15, I16, I21, I22, I23, I24
- Kompetensi Kepribadian (*responsiveness*): I6, I17
- Kompetensi Profesional (*reliability*): I7, I20
- Aspek aksesibilitas dan sarana Prasarana (*tangible*): I18, I19.

Tabel 2 menjelaskan masing-masing pertanyaan beserta klasifikasinya:

Tabel 2. Pertanyaan Variabel Prediktor

No	Kompetensi	Pertanyaan
1	Sosial ( <i>empathy</i> )	Dosen mampu berkomunikasi dengan baik dengan mahasiswa.
2	Pedagogik ( <i>assurance</i> )	Dosen mampu menciptakan suasana di kelas dengan membangkitkan semangat belajar dan mendorong minat diskusi.
3	Sosial ( <i>empathy</i> )	Dosen menghargai pendapat mahasiswa.
4	Sosial ( <i>empathy</i> )	Dosen menghargai keragaman mahasiswa dari suku atau agama ataupun adat istiadat.
5	Sosial ( <i>empathy</i> )	Dosen tanggap terhadap permasalahan yang terjadi dalam pembelajaran di kelas.
6	Kepribadian ( <i>responsiveness</i> )	Kearifan dan keadilan dalam mengambil keputusan dan menyelesaikan persoalan.
7	Profesional ( <i>reliability</i> )	Kedalaman dan keluasan dosen dalam membahas contoh kasus yang terkait bidang topik bahasan dan penggunaan hasil penelitian yang terkait.
8	Pedagogik ( <i>assurance</i> )	Kejelasan dosen dalam menyampaikan tujuan pembelajaran dan materi serta jawaban terhadap pertanyaan di kelas.
9	Profesional ( <i>reliability</i> )	Kemampuan dosen dalam menguasai wawasan dan mengimplementasikan mata kuliah yang diberikan.
10	Profesional ( <i>reliability</i> )	Kemampuan dosen dalam menjawab pertanyaan yang diajukan oleh mahasiswa.
11	Profesional ( <i>reliability</i> )	Kemampuan dosen dalam menjelaskan materi perkuliahan secara sistematis.
12	Pedagogik ( <i>assurance</i> )	Keragaman metode pembelajaran yang disampaikan dosen dalam diskusi dan tanya jawab.
13	Pedagogik ( <i>assurance</i> )	Kesesuaian soal ujian dengan materi ajar.
14	Pedagogik ( <i>assurance</i> )	Ketepatan waktu kehadiran dosen dan mengajar sesuai waktu yang terjadwal.
15	Pedagogik ( <i>assurance</i> )	Pemberian tugas terstruktur seperti paper atau rangkuman atau latihan soal dan pemecahan masalah.
16	Pedagogik ( <i>assurance</i> )	Rencana materi dan tujuan kuliah disusun sistematis dan terorganisir serta disampaikan di awal perkuliahan.
17	Kepribadian ( <i>responsiveness</i> )	Santun dalam kata dan tindakan serta dapat dijadikan teladan dalam bersikap.
18	Aksesibilitas dan Sarpras ( <i>tangible</i> )	Sarana dan prasarana pembelajaran selalu tersedia dalam kondisi baik dan layak guna.
19	Aksesibilitas dan Sarpras ( <i>tangible</i> )	Sarana dan prasarana selalu tersedia untuk digunakan dalam jumlah cukup.
20	Profesional ( <i>reliability</i> )	Semangat dosen dalam memberikan kuliah serta

No	Kompetensi	Pertanyaan
21	Pedagogik ( <i>assurance</i> )	menjelaskan kembali materi yang tidak dipahami dan memberikan motivasi minat belajar. Sumber belajar yang disampaikan dosen dalam perkuliahan mengarah sesuai atas kasus lapangan atau pengembangan sendiri.
22	Pedagogik ( <i>assurance</i> )	Tersedianya bahan ajar untuk perkuliahan offline dan online.
23	Pedagogik ( <i>assurance</i> )	Tersedianya kontrak kuliah sebelum perkuliahan di mulai.
24	Pedagogik ( <i>assurance</i> )	Tersedianya Rencana Pembelajaran Semester di awal perkuliahan.

### 2.1.2 Preprocessing

Sebelum masuk ke tahap pemodelan, proses *preprocessing* data menjadi langkah penting untuk memastikan kualitas data yang akan digunakan. Tahap ini bertujuan untuk memverifikasi bahwa data yang dianalisis benar-benar bersih, konsisten, dan layak diproses oleh sistem pemodelan.

Dalam penelitian ini, proses diawali dengan pengambilan data awal ke dalam pemrograman R. Tahapan pertama dalam *preprocessing* yang dilakukan adalah pemeriksaan terhadap nilai kosong (*missing values*). Langkah ini sangat penting untuk memastikan bahwa data yang digunakan bersih dari kesalahan struktural, karena keberadaan nilai kosong dapat mengakibatkan galat dalam proses pemodelan atau menurunkan akurasi hasil prediksi. Setelah dipastikan bahwa data tidak mengandung nilai kosong, proses dilanjutkan dengan pemisahan antara variabel fitur (X) dan label (Y). Y ditetapkan sebagai variabel target yang akan diprediksi, sementara kolom-kolom lainnya digunakan sebagai variabel input.

Tahapan selanjutnya dalam *preprocessing* adalah identifikasi nilai-nilai ekstrem (*outlier*) melalui visualisasi *boxplot* terhadap seluruh variabel numerik. Deteksi dini terhadap *outlier* penting dilakukan karena nilai-nilai ekstrem dapat mempengaruhi distribusi data dan mengganggu kinerja model prediksi.

Efektivitas tahapan ini juga diperkuat oleh temuan dari Nur et al. (2023), yang menunjukkan bahwa penerapan teknik *preprocessing* secara cermat terutama dalam mendeteksi dan memperbaiki *missing value* dan *outlier* secara signifikan mampu meningkatkan akurasi prediksi model *Random Forest Regression* dalam konteks pertanian, khususnya prediksi hasil panen padi [15].

### 2.1.3 Eksplorasi Data

Sebelum melangkah ke tahap pemodelan, dilakukan eksplorasi awal untuk memahami karakteristik data yang akan dianalisis. Visualisasi data dilakukan menggunakan histogram dari nilai *Mean* serta *Varians* tiap variabel prediktor (I1–I24). Histogram *Mean* dan *Varians* memberikan gambaran tentang kecenderungan nilai tengah serta sejauh mana data tersebar untuk masing-masing indikator penilaian.

Langkah ini penting untuk memastikan bahwa data memiliki pola distribusi yang wajar dan tidak ada

penyimpangan ekstrem yang bisa mengganggu proses prediksi.

## 2.2 Pemodelan dengan *Random Forest Regression*

### 2.2.1 Pemisahan Data dan Seleksi Fitur

Pada tahap awal pemodelan, dataset dibagi ke dalam dua kelompok utama: 70% digunakan sebagai data latih (*training set*) dan sisanya 30% sebagai data uji (*testing set*). Pembagian ini dilakukan secara proporsional dalam perangkat lunak R, dengan mempertimbangkan distribusi target agar tetap seimbang. Tujuannya adalah untuk memberikan ruang bagi model belajar dari data yang tersedia, sekaligus menguji kemampuannya dalam memprediksi data yang belum dikenalnya secara obyektif. Teknik pembagian seperti ini banyak digunakan dalam penelitian prediktif berbasis *Random Forest Regression* karena dinilai efektif menjaga generalisasi model [15], [16].

Setelah data dibagi, langkah berikutnya adalah melakukan seleksi fitur guna menyaring variabel-variabel yang paling relevan. Dalam penelitian ini digunakan pendekatan *Recursive Feature Elimination* (RFE), sebuah teknik yang bekerja secara bertahap dengan mengeliminasi fitur yang memiliki kontribusi paling rendah terhadap kualitas prediksi model. Untuk meningkatkan validitas hasil seleksi, proses RFE ini dikombinasikan dengan *lima-fold cross-validation*. Strategi validasi silang ini membantu memastikan bahwa pemilihan fitur tidak bias terhadap *subset* data tertentu, sehingga model yang dihasilkan lebih stabil dan akurat.

Efektivitas metode ini didukung oleh penelitian Wiguna et al. (2025), yang membuktikan bahwa penggunaan RFE dengan validasi silang lima lipat mampu meningkatkan performa model *Random Forest Regression* melalui pemilihan atribut yang benar-benar relevan dalam konteks prediksi [13].

### 2.2.2 Pemodelan dan Tuning Parameter

Pada tahap ini, model prediksi dibangun menggunakan algoritma *Random Forest Regression* (RFR), yang terkenal karena kemampuannya menangani data kompleks dan memberikan hasil yang akurat. Untuk memastikan performa model berada pada tingkat optimal, dua parameter utama yang memengaruhi struktur dan kemampuan model disesuaikan, yaitu *mtry* (jumlah variabel yang dipertimbangkan saat setiap pohon melakukan *split*) dan *ntree* (jumlah total pohon dalam hutan).

Proses penyesuaian parameter ini dilakukan menggunakan teknik *grid search*, yaitu pendekatan sistematis yang menguji berbagai kombinasi nilai *mtry* dan *ntree*. Tujuan dari metode ini adalah untuk menemukan kombinasi parameter yang memberikan nilai kesalahan prediksi terkecil, yang dalam hal ini diukur menggunakan *Root Mean Square Error* (RMSE). Nilai RMSE yang lebih rendah menunjukkan bahwa prediksi model semakin mendekati nilai aktual, menandakan performa yang semakin baik.

Penggunaan *grid search* dalam tuning parameter bukanlah hal baru. Han dan Kim (2021) menunjukkan bahwa proses tuning parameter memiliki pengaruh signifikan terhadap akurasi model prediktif. Dalam penelitiannya,

mereka membuktikan bahwa optimasi nilai *mtry* dan *nree* dapat menurunkan kesalahan prediksi secara substansial dan meningkatkan reliabilitas hasil prediksi [16].

### 2.3 Evaluasi Model

Setelah proses pelatihan model *Random Forest Regression* (RFR) selesai, langkah berikutnya adalah mengevaluasi seberapa baik model tersebut mampu melakukan prediksi terhadap data yang belum pernah dilihat sebelumnya, yaitu data uji. Evaluasi dilakukan melalui dua pendekatan: kuantitatif dan visual, dengan fokus utama pada metrik evaluasi regresi yang telah diakui luas dalam penelitian sebelumnya.

#### 2.3.1 Evaluasi Kuantitatif

Evaluasi objektif terhadap kinerja model dilakukan dengan menggunakan tiga metrik utama dalam analisis regresi, yakni *Root Mean Square Error* (RMSE), *Mean Absolute Error* (MAE), serta Koefisien Determinasi ( $R^2$ ). Ketiga metrik ini saling melengkapi dan memberikan gambaran menyeluruh terhadap keakuratan serta keandalan model prediktif.

##### 2.3.1.1 RMSE

RMSE mengukur rata-rata akar kuadrat dari selisih antara nilai aktual dengan nilai prediksi. Metrik ini sangat sensitif terhadap pencilan, sehingga memberikan penalti yang lebih besar untuk kesalahan prediksi yang ekstrem. Nilai RMSE yang rendah mengindikasikan bahwa model mampu memprediksi dengan baik dan hasilnya mendekati nilai sebenarnya. Wiguna et al. mencatat bahwa RMSE efektif dalam mengidentifikasi seberapa konsisten model dalam memprediksi nilai kontinyu pada penelitian regresi [13].

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2} \quad (1)$$

##### 2.3.1.2 Mean Absolute Error (MAE)

MAE mengukur rata-rata dari nilai absolut selisih antara nilai aktual dan nilai prediksi. Berbeda dari RMSE, MAE lebih tahan terhadap pengaruh pencilan, sehingga memberikan gambaran yang stabil terhadap rata-rata kesalahan prediksi. Dalam konteks ini, MAE sangat berguna untuk menilai ketepatan prediksi secara umum tanpa terlalu terpengaruh oleh nilai ekstrem. Seperti dijelaskan oleh Ardianto et al., MAE cocok digunakan sebagai pelengkap RMSE dalam mengevaluasi model regresi [14].

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |Y_i - \hat{Y}_i| \quad (2)$$

##### 2.3.1.3 Koefisien Determinasi ( $R^2$ )

$R^2$  mengukur sejauh mana *Variansi* dari variabel target ( $Y$ ) dapat dijelaskan oleh model. Nilai  $R^2$  berada dalam rentang 0 hingga 1, di mana nilai mendekati 1 menunjukkan bahwa model memiliki kemampuan prediksi yang sangat tinggi. Menurut Ardianto et al., nilai  $R^2$  yang tinggi mencerminkan bahwa model berhasil menangkap pola

dalam data dengan baik, menjadikannya alat yang kuat untuk evaluasi performa dalam regresi [14].

$$R^2 = 1 - \frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{\sum_{i=1}^n (Y_i - \bar{Y})^2} \quad (3)$$

Melalui kombinasi ketiga metrik ini, evaluasi model dilakukan secara komprehensif, memastikan bahwa model tidak hanya akurat secara rata-rata, tetapi juga konsisten dan dapat diandalkan dalam menangani variasi data nyata.

#### 2.3.2 Evaluasi Visual (*Scatterplot* Prediksi vs Aktual)

Sebagai pelengkap dari evaluasi kuantitatif, pendekatan visual juga digunakan untuk memberikan gambaran intuitif terhadap performa model. Salah satu metode visualisasi yang paling umum dalam konteks regresi adalah *scatterplot* antara nilai prediksi ( $\hat{Y}_i$ ) dan nilai aktual ( $Y_i$ ).

Visualisasi ini menggambarkan sebaran titik hasil prediksi terhadap nilai aktual pada bidang dua dimensi, di mana garis imajiner  $y = x$  digunakan sebagai acuan ideal yaitu saat prediksi sepenuhnya sama dengan nilai aktual. Semakin dekat titik-titik data terhadap garis ini, semakin tinggi tingkat akurasi dan konsistensi model. Selain menilai kedekatan prediksi dengan observasi, *scatterplot* juga berfungsi penting dalam mengidentifikasi pola kesalahan sistematis serta *outlier* yang mungkin memengaruhi performa model secara keseluruhan. Pendekatan ini terbukti efektif dalam evaluasi model regresi sebagaimana diterapkan oleh Pratama dan Nugroho dalam penelitian mereka terkait prediksi harga rumah menggunakan *Random Forest Regression* [17].

### 2.4 Analisis *Feature importance*

Setelah model *Random Forest Regression* (RFR) berhasil dibangun dan divalidasi, langkah lanjutan yang krusial adalah memahami kontribusi masing-masing variabel prediktor terhadap kinerja model. Tujuan dari tahap ini bukan hanya menilai akurasi prediksi, tetapi juga mendapatkan pemahaman yang lebih dalam mengenai variabel mana yang paling berpengaruh dalam membentuk keputusan model.

Salah satu pendekatan yang digunakan adalah analisis *feature importance* dengan metrik *%IncMSE* (*Percent Increase in Mean Squared Error*). Metrik ini bekerja berdasarkan prinsip *permutation importance*, yaitu dengan mengacak nilai dari setiap variabel prediktor dan mengamati sejauh mana kesalahan prediksi (MSE) meningkat. Semakin besar kenaikan MSE akibat pengacakan suatu variabel, maka semakin penting peran variabel tersebut dalam menjaga akurasi model.

Keunggulan metode ini adalah kemampuannya dalam:

- Mengidentifikasi fitur-fitur paling esensial dalam proses prediksi;
- Menyaring fitur yang kurang relevan untuk menghindari kompleksitas model yang berlebihan;
- Memberikan wawasan berbasis data yang dapat mendukung pengambilan kebijakan yang lebih tepat sasaran.

Tidak seperti teknik regresi tradisional, metode ini tidak mengasumsikan hubungan linier atau korelasi antar variabel, sehingga sangat sesuai untuk data yang kompleks dan berdimensi tinggi.

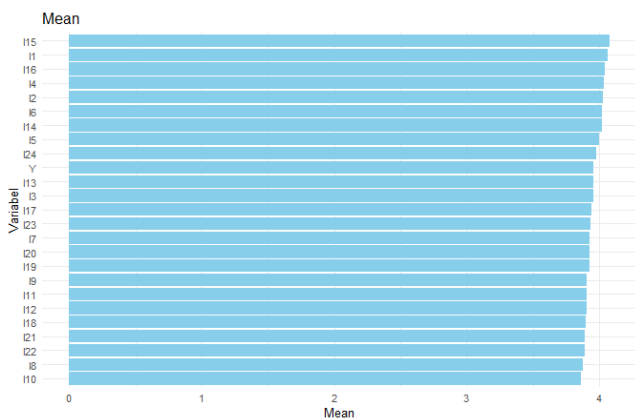
Dalam penelitian Wiguna et al. (2025), metode %IncMSE terbukti efektif dalam mengungkap fitur-fitur paling berpengaruh dalam prediksi harga smartphone. Hasilnya menunjukkan bahwa pendekatan ini memberikan interpretasi yang kuat terhadap struktur data dan dapat diandalkan untuk memahami elemen-elemen yang paling menentukan dalam proses prediktif [13].

### 3. Hasil dan Pembahasan

#### 3.1 Statistik Deskriptif

Sebelum melangkah ke tahap pemodelan prediktif, peneliti melakukan eksplorasi awal terhadap data untuk memahami lebih dalam bagaimana mahasiswa menilai dosen. Eksplorasi ini bertujuan untuk menggali distribusi skor, mengidentifikasi pola umum, serta mendeteksi potensi anomali atau kecenderungan nilai ekstrem yang dapat memengaruhi akurasi model prediktif.

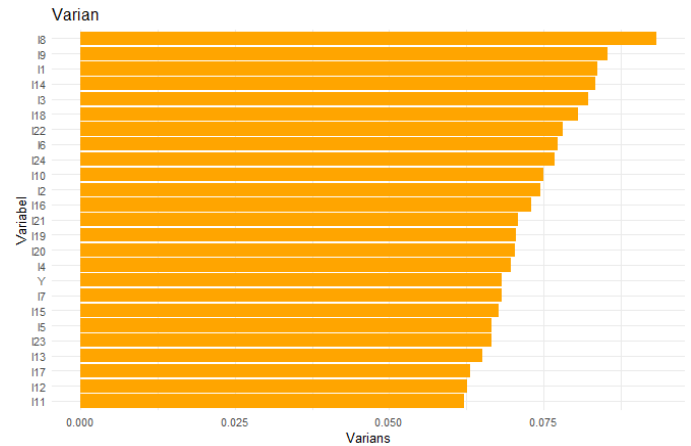
##### 3.1.1 Grafik Mean Semua Variabel



Gambar 2. Grafik Mean Semua Variabel

Gambar 2 menampilkan rata-rata nilai dari 24 indikator penilaian dosen (I1–I24) dan satu variabel target Y yang merepresentasikan tingkat kepuasan mahasiswa secara keseluruhan. Mayoritas indikator memiliki nilai *Mean* di atas 3, menunjukkan persepsi positif mahasiswa terhadap kinerja dosen. Tiga indikator tertinggi adalah I15 (tugas terstruktur), I1 (komunikasi), dan I6 (kebijaksanaan), mencerminkan aspek yang paling diapresiasi. Sementara itu, indikator I10, I8, dan I11 memiliki nilai rata-rata lebih rendah, namun masih dalam kategori baik. Nilai *Mean* variabel Y berada di kisaran tengah, mengindikasikan kemampuannya sebagai representasi agregat dari seluruh indikator. Distribusi *Mean* yang stabil dan tidak ekstrem memperkuat fondasi model prediktif dan mendukung proses seleksi fitur melalui metode *Recursive Feature Elimination* (RFE). Pendekatan eksplorasi statistik ini juga digunakan oleh Wiguna et al. (2025) sebagai langkah awal dalam membangun model prediksi yang optimal [13].

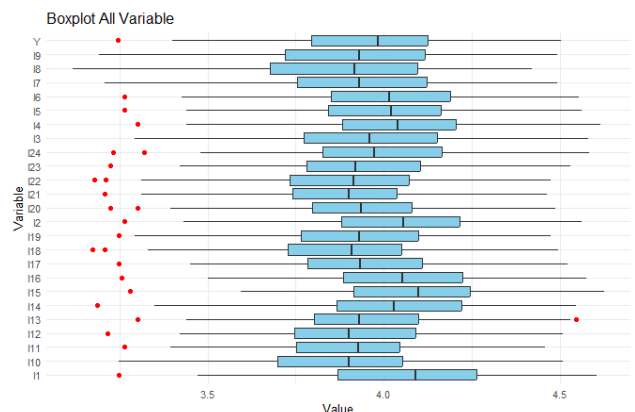
##### 3.1.2 Grafik Varians Semua Variabel



Gambar 3. Grafik Varians Semua Variabel

Gambar 3 menampilkan nilai varian dari 24 indikator penilaian dosen (I1–I24) untuk mengukur seberapa besar penyebaran penilaian mahasiswa terhadap masing-masing indikator. Varian yang tinggi mencerminkan keragaman persepsi mahasiswa, sehingga indikator tersebut berpotensi besar dalam membedakan karakteristik penilaian dan memperkaya model prediktif. Indikator seperti I8 (kejelasan tujuan pembelajaran), I9 (pemahaman materi), I1 (komunikasi), dan I4 (penghargaan keberagaman) menunjukkan varian tinggi, yang menandakan persepsi mahasiswa cukup bervariasi terhadap aspek tersebut. Sebaliknya, I11 (penggunaan media), I12 (keragaman metode), dan I17 (sikap dosen) memiliki varian rendah, menunjukkan konsistensi penilaian yang tinggi, namun kontribusinya terhadap pembeda model mungkin terbatas. Analisis ini menjadi tahap awal penting dalam proses seleksi fitur menggunakan metode *Recursive Feature Elimination* (RFE), guna mengidentifikasi variabel yang paling signifikan dalam meningkatkan akurasi prediksi tingkat kepuasan mahasiswa.

##### 3.1.3 Boxplot Seluruh Variabel Penilaian



Gambar 4. Grafik Boxplot Semua Variabel

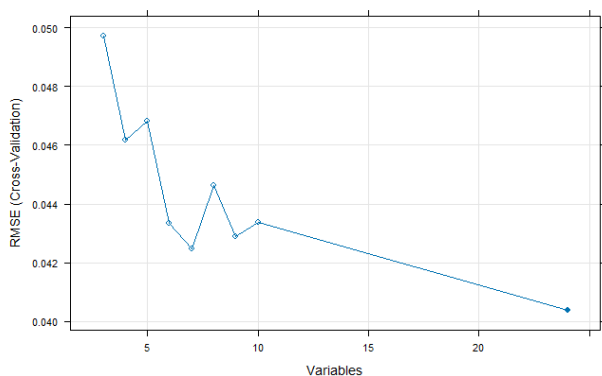
Gambar 4 menyajikan *boxplot* dari 24 indikator penilaian dosen (I1–I24) dan variabel target Y untuk menunjukkan sebaran data, posisi median, dan keberadaan *outlier*. Sebagian besar indikator memiliki median

mendekati 4, menandakan penilaian mahasiswa terhadap kinerja dosen tergolong baik. Indikator dengan rentang interkuartil (IQR) sempit menunjukkan penilaian yang konsisten, sedangkan indikator seperti I8 (kejelasan penyampaian tujuan pembelajaran) dan I15 (pemberian tugas terstruktur) memiliki IQR yang lebih lebar, menandakan variasi persepsi yang tinggi di kalangan mahasiswa. Beberapa indikator juga menunjukkan adanya *outlier*, yaitu penilaian ekstrem yang dapat memengaruhi stabilitas model prediktif. Visualisasi ini menjadi dasar penting dalam menilai kualitas data sebelum dilakukan pemodelan, dan pendekatan serupa telah digunakan oleh Wiguna et al. (2025) dalam membangun model prediksi menggunakan *Random Forest Regression* [13].

### 3.2 Pemodelan *Random Forest Regression*

Dalam tahapan pemodelan prediktif, penelitian ini memanfaatkan algoritma *Random Forest Regression* (RFR) dengan pendekatan yang menyeluruh. Proses dimulai dari pemisahan dataset menjadi dua bagian utama, yaitu data latih (70%) dan data uji (30%). Pemisahan ini dilakukan agar model dapat belajar dari sebagian besar data yang tersedia, lalu diuji pada data yang belum pernah dikenali sebelumnya. Strategi ini tidak hanya memungkinkan model mengenali pola yang ada dalam data, tetapi juga mengukur kemampuannya dalam melakukan prediksi yang akurat terhadap data baru, sejauh mana model dapat memahami struktur data secara menyeluruh.

Langkah awal setelah pembagian data adalah penerapan teknik *Recursive Feature Elimination* (RFE) yang dikombinasikan dengan *5-fold cross-validation*. Teknik ini diterapkan untuk memilih fitur-fitur (variabel variabel independen) yang paling relevan dan berkontribusi besar terhadap prediksi. RFE bekerja dengan cara menyusun model secara bertahap dan mengeliminasi variabel independen yang memiliki pengaruh terendah terhadap akurasi. Kombinasi dengan *5-fold cross-validation* membantu memastikan bahwa proses seleksi fitur tidak hanya berlaku untuk satu subset data, tetapi tervalidasi di berbagai bagian data latih. Hasil dari proses ini divisualisasikan dalam Gambar 5.



Gambar 5. Grafik RMSE Model RFE Berdasarkan Jumlah Variabel Independen

Sebagaimana ditampilkan pada Gambar 5, terlihat bahwa semakin banyak variabel independen yang dimasukkan ke dalam model, semakin kecil nilai RMSE

yang dihasilkan. Tren ini menunjukkan bahwa setiap penambahan variabel memberikan kontribusi terhadap peningkatan akurasi model, hingga pada titik optimal ketika semua variabel digunakan.

Temuan ini selaras dengan hasil penelitian Han dan Kim [16], yang menyatakan bahwa pemilihan subset fitur yang tepat secara signifikan dapat menurunkan nilai RMSE dalam model *Random Forest Regression*.

### 3.3 Evaluasi Model

#### 3.3.1 Evaluasi Kuantitatif

Setelah mendapatkan kombinasi fitur terbaik, proses dilanjutkan dengan tuning parameter *mtry* dan *ntree* menggunakan teknik *grid search*. Nilai *mtry* diatur mulai dari 1 hingga 24, sedangkan *ntree* diuji pada 100, 300, dan 500 pohon. Tujuan dari tuning ini adalah mengidentifikasi konfigurasi parameter yang mampu menghasilkan galat prediksi terkecil (RMSE) pada data uji.

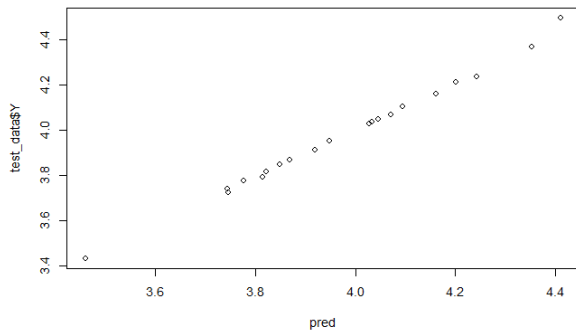
```
> print(eval)
      RMSE  Rsquared      MAE
0.02228402 0.99594877 0.01184139

Call:
  randomForest(formula = as.formula(paste("Y
~", paste(selected_vars, collapse =
"+"))), data = train_data, mtry = best$mtry,
  ntree = best$ntree, importance = TRUE)
  Type of random forest:
  regression
                        Number of trees: 300
No. of variables tried at each split: 1
      Mean of squared residuals:
0.001934501
                        % Var explained: 97.26
```

Berdasarkan hasil *grid search*, diperoleh konfigurasi terbaik pada *mtry* = 1 dan *ntree* = 300, yang digunakan untuk membentuk model akhir. Hasil evaluasi menunjukkan bahwa model *Random Forest Regression* (RFR) memiliki kinerja yang sangat baik dalam memprediksi tingkat kepuasan mahasiswa terhadap kinerja dosen. Nilai *Root Mean Square Error* (RMSE) sebesar 0,0222 dan *Mean Absolute Error* (MAE) sebesar 0,0118 mengindikasikan tingkat kesalahan prediksi yang sangat kecil. Selain itu, nilai Koefisien Determinasi ( $R^2$ ) sebesar 0,9959 mengindikasikan bahwa model mampu menjelaskan sekitar 99,59% variasi yang terdapat pada variabel target (Y).

#### 3.3.2 Evaluasi Visual (*Scatterplot* Prediksi vs Aktual/Observasi)

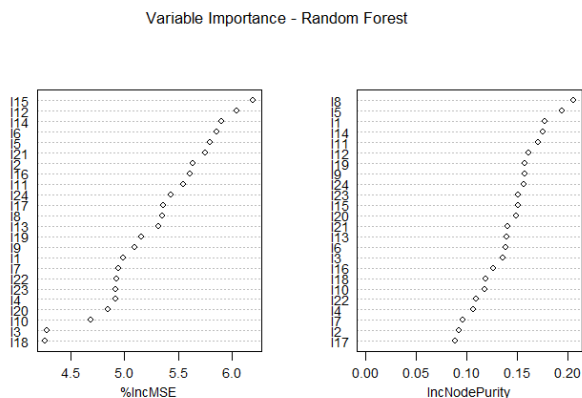
Untuk memahami sejauh mana prediksi model mendekati nilai aktual, dilakukan evaluasi visual menggunakan *scatterplot* yang memetakan hubungan antara hasil prediksi ( $\hat{Y}$ ) dan nilai observasi (Y), sebagaimana ditampilkan pada Gambar 6.



Gambar 6. Scatterplot Antara Prediksi ( $\hat{Y}$ ) dan  $Y$  Observasi

Sebagian besar titik data tampak tersebar mengikuti pola diagonal dari kiri bawah ke kanan atas, mendekati garis imajiner  $y = x$ . Pola ini menunjukkan bahwa hasil prediksi model sangat mendekati nilai aktual/observasi, sehingga dapat disimpulkan bahwa akurasi model cukup tinggi. Tidak ditemukan *outlier* ekstrem ataupun pola deviasi sistematis, yang menandakan bahwa model bekerja secara konsisten dan stabil dalam menghasilkan estimasi nilai kepuasan mahasiswa. Visualisasi ini juga memperkuat hasil evaluasi numerik sebelumnya, serta memberikan bukti tambahan bahwa model mampu merepresentasikan hubungan yang kuat antara variabel input dan target.

### 3.4 Analisis Feature importance



Gambar 7. Variabel Importance Model RF

Gambar 7 menampilkan hasil analisis *feature importance* berdasarkan metrik %IncMSE (*Percent Increase in Mean Squared Error*). Metrik ini menunjukkan sejauh mana tingkat kesalahan prediksi (MSE) meningkat ketika nilai suatu variabel diacak secara acak. Semakin besar peningkatan MSE akibat pengacakan, semakin penting peran variabel tersebut dalam menjaga akurasi model.

Hasil analisis menunjukkan bahwa tiga variabel dengan pengaruh paling besar terhadap performa model adalah I15, I12, dan I14. Ketiganya menyebabkan lonjakan MSE yang paling tinggi saat nilainya diacak, yang berarti ketiganya merupakan prediktor utama dalam menentukan tingkat kepuasan mahasiswa. Secara khusus:

- I15 berkaitan dengan pemberian tugas terstruktur seperti makalah, latihan soal, atau pemecahan masalah.

- I12 menggambarkan keragaman metode pembelajaran yang digunakan dosen dalam sesi diskusi dan tanya jawab.
- I14 merujuk pada ketepatan waktu kehadiran dosen serta kesesuaian jadwal pengajaran.

Temuan ini memperkuat pentingnya dimensi pedagogik dalam membentuk pengalaman belajar mahasiswa yang positif.

## 4 Kesimpulan

Penelitian ini berhasil merancang dan menerapkan model prediktif berbasis algoritma *Random Forest Regression* untuk mengestimasi tingkat kepuasan mahasiswa terhadap dosen. Model dikembangkan dengan memanfaatkan 24 indikator penilaian yang merepresentasikan aspek kompetensi pedagogik, profesional, kepribadian, sosial, serta dukungan sarana dan prasarana. Evaluasi yang dilakukan mengungkapkan bahwa model menunjukkan kinerja prediksi yang sangat tinggi, dengan nilai RMSE sebesar 0,0222, MAE sebesar 0,0118, dan koefisien determinasi ( $R^2$ ) mencapai 0,9959. Hasil ini mencerminkan kemampuan model dalam menjelaskan hampir seluruh variasi yang terdapat pada data target, sehingga dapat dianggap sebagai alat prediktif yang sangat dapat diandalkan dalam analisis berbasis data.

Melalui analisis *feature importance*, penelitian ini juga berhasil mengidentifikasi tiga indikator paling berpengaruh terhadap kepuasan mahasiswa, yaitu: I15 (pemberian tugas terstruktur), I12 (keragaman metode pembelajaran), I14 (ketepatan waktu kehadiran dosen). Temuan ini menekankan pentingnya kejelasan struktur pembelajaran serta kedisiplinan dosen sebagai faktor utama dalam membentuk persepsi positif dari mahasiswa.

Selain keberhasilan dalam membangun model prediktif, penelitian ini turut memberikan kontribusi pada aspek metodologis melalui penerapan tuning parameter (*mtry* dan *ntree*) dengan teknik *grid search* dan validasi silang. Pendekatan ini terbukti meningkatkan akurasi dan stabilitas model, serta mengatasi kekurangan dari penelitian sebelumnya yang cenderung menggunakan konfigurasi parameter secara *default*. Dengan hasil yang diperoleh, model ini memiliki potensi untuk dijadikan dasar dalam merumuskan strategi peningkatan mutu pembelajaran yang berkelanjutan di lingkungan Politeknik Negeri Cilacap, sekaligus menjadi rujukan dalam pengembangan sistem evaluasi dosen berbasis kecerdasan buatan.

## Ucapan Terimakasih

Terimakasih kami sampaikan kepada Politeknik Negeri Cilacap khususnya Pusat Pengembangan Pembelajaran dan Penjaminan Mutu, yang telah menyediakan data yang diperlukan dalam penelitian..

## Daftar Pustaka

- [1] G. M. Aji, K. Murkatik, E. Harapan, and D. Wardiah, "The Influence of Professional and Pedagogic Competence on Teacher's Performance," *Journal of Social Work and Science Education*, vol. 1, no. 1, pp. 58–69, 2020..
- [2] A. Almasri, S. Jamil, and N. S. Ahmad, "Student satisfaction prediction in online learning using Random Forest classifier," *Int. J. Adv. Comput. Sci. Appl. (IJACSA)*, vol. 13, no. 4, pp.

- 565–571, 2022.
- [3] R. Widayati, A. F. Pratama, and M. A. Nugraha, “Prediction of student loyalty using *machine learning*: Random Forest approach,” *J. Teknol. Inform. dan Pendidik.*, vol. 15, no. 1, pp. 45–51, 2022.
- [4] K. Nachouki, M. Laskri, and S. Ben Amor, “Predicting student academic performance using Random Forest: Evidence from Algerian higher education,” *Educ. Inf. Technol.*, vol. 26, no. 3, pp. 3211–3230, 2021.
- [5] D. Riyanto, S. Susanto, and Y. Prabowo, “Analysis of student satisfaction using various classification algorithms: A case study,” *J. Inf. Syst. Eng. Bus. Intell.*, vol. 7, no. 2, pp. 120–130, 2021.
- [6] Sufina dan L. Wati, “Implementasi Algoritma Random Forest Pada Sistem Penilaian Kinerja Dosen,” *JAISE*, vol. 5, no. 2, pp. 790–798, 2025.
- [7] Tiftazani, A. Tenriawaru, and G. A. Rahman, “Analisis Klasifikasi Kepuasan Mahasiswa terhadap Penyelenggaraan Pelayanan Akademik FMIPA Universitas Halu Oleo Menggunakan Algoritma Random Forest,” *AnoaTIK: J. Teknol. Inf. dan Komput.*, vol. 2, no. 2, pp. 77–83, Dec. 2024.
- [8] S. Rani and A. Rakhmawati, “Tingkat Kepuasan Mahasiswa atas Kinerja Dosen pada Sistem Perkuliahan Online Selama Masa Pandemi,” *Strategi*, vol. 12, no. 1, pp. 13–28, Apr. 2022, doi: 10.52333/strategi.v12i1.365.
- [9] S. T. Selviani, I.W. Karta, dan S. Wilian, “Pengaruh Kompetensi Pedagogik terhadap Kinerja Akademik Dosen,” *Jurnal Ilmu Sosial dan Pendidikan*, vol. 4, no. 4, 2020.
- [10] A. E. Suryanto, “Student Satisfaction With Teaching Performance, Academic Administration Services, And Learning Facilities,” *Khidmatuna: J. Res. Community Serv.*, vol. 1, no. 2, pp. 19–29, Mar. 2023, doi: 10.58330/khidmatuna.v2i1.230.
- [11] P. Ramadani, G. W. Nurcahyo, dan B. Hendrik, “Penerapan Metode Naïve Bayes dalam Memprediksi Kepuasan Mahasiswa Terhadap Cara Pengajaran Dosen,” *KESATRIA*, vol. 5, no. 2, 2024.
- [12] P.P.P.A.N.W. Fikrul Ilmi, et al., “Analisis Model Kepuasan Mahasiswa terhadap Cara Pengajaran Dosen Menggunakan Algoritma C4.5,” *JIKO*, vol. 6, no. 1, 2022.
- [13] S. Wiguna, S. P. A. Alkadri, dan Istikoma, “Prediksi Harga Smartphone Berdasarkan Fitur Smartphone Dengan *Random Forest Regression*,” *Kohesi: Jurnal Multidisiplin Saintek*, vol. 6, no. 9, pp. 1–7, Jan. 2025.
- [14] A. Ardianto, A. B. Raharjo, dan D. Purwitasari, “*Random Forest Regression* Untuk Prediksi Produksi Daya Pembangkit Listrik Tenaga Surya,” *BRILIANT: Jurnal Riset dan Konseptual*, vol. 7, no. 4, pp. 1058–1075, Nov. 2022, doi: 10.28926/briliant.v7i4.1036.
- [15] N. Nur, F. Wajidi, Sulfayanti, dan Wildayani, “Implementasi Algoritma *Random Forest Regression* untuk Memprediksi Hasil Panen Padi di Desa Minanga,” *Jurnal Komputer Terapan*, vol. 9, no. 1, pp. 10–18, 2023.
- [16] M. Han and J. Kim, “Optimal Feature Set Size in *Random Forest Regression* Using *Grid search* with RMSE Evaluation,” *Applied Sciences*, vol. 11, no. 7, pp. 3101–3112, 2021, doi: 10.3390/app11073101.
- [17] R. A. Pratama dan F. S. Nugroho, “Perbandingan Performa Algoritma Linear Regresi dan Random Forest dalam Memprediksi Harga Rumah,” *Jurnal Informatika dan Teknik Komputer (JINTEKS)*, vol. 8, no. 2, pp. 78–86, 2022.