

Machine Learning Analysis Using the Hybrid K-Means and C4.5 Model in Determining Student Readiness for Educational Field Practice

Nilam Agsera ¹⁾; Musli Yanto ²⁾; Syafri Arlis ³⁾

¹⁾ Department of Informatics Engineering Education, Faculty of Engineering, Universitas Negeri Padang, Padang

^{2,3)} Department of Informatics Engineering, Faculty of Comomputer Science, Universitas Putra Indonesia YPTK, Padang Sumatera Barat

Email: ¹⁾ ppg.nilamagsera96330@program.belajar.id

How to Cite :

Agsera, N., Yanto, M., Arlis, S. (2026). Machine Learning Analysis Using the Hybrid K-Means and C4.5 Model in Determining Student Readiness for Educational Field Practice Jurnal Media Computer Science, 5(4)

ARTICLE HISTORY

Received [03 Agustus 2026]

Revised [25 Oktober 2026]

Accepted [27 Oktober 2026]

KEYWORDS

Machine Learning; K-Means; C4.5; Student Readiness; Educational Field Practice.

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



ABSTRAK

Evaluasi kesiapan mahasiswa Praktik Lapangan Kependidikan (PLK) selama ini masih bergantung pada penilaian subjektif dosen pembimbing dan guru pamong sehingga berisiko menimbulkan masalah ketidaktepatan hasil yang diperoleh. Model tersebut masih konvensional dengan analisis berbasis pendekatan manual. Berdasarkan temuan tersebut maka, penelitian ini bertujuan membangun model hybrid K-Means dan C4.5 untuk menentukan kategori kesiapan mahasiswa PLK di Universitas Negeri Padang. K-Means diterapkan untuk mengelompokkan data penilaian menjadi tiga kategori kesiapan, kemudian label kluster yang terbentuk dijadikan target analisis. C4.5 digunakan untuk menyajikan rule dalam pohon keputusan untuk penentuan hasil. Dataset berjumlah 3.578 rekaman nilai mahasiswa yang diperoleh dari Pusat PPL Universitas Negeri Padang periode Juli-Desember 2025, mencakup sepuluh variabel indikator penilaian dari observasi sekolah mitra hingga laporan akhir praktik. Pengklusteran K-Means mengelompokkan 2.619 mahasiswa (73,20%) ke dalam kategori Siap, 936 mahasiswa (26,16%) kategori Cukup Siap, dan 23 mahasiswa (0,64%) kategori Perlu Pendampingan. Model C4.5 memilih indikator media pembelajaran (X4) sebagai simpul akar (root node) dan mencapai akurasi keseluruhan sebesar 92,32%, presisi terbobot 92,33%, recall 92,32%, dan F1-score 92,33%. Temuan ini menunjukkan bahwa pendekatan hybrid K-Means-C4.5 dapat menyediakan instrumen yang objektif, transparan, dan berbasis data untuk membantu pengelola PLK dalam penilaian akhir PLK dan menetapkan status kesiapan mahasiswa setelah menyelesaikan seluruh rangkaian praktik di sekolah mitra.

ABSTRACT

The evaluation of student readiness for Educational Field Practice (PLK) has conventionally relied on the subjective assessments of supervising lecturers and mentor teachers, posing a risk of inaccurate outcomes. This traditional framework continues to depend on manual, non-standardized analytical approaches. To address these limitations, this study aims to develop a hybrid K-Means and C4.5 machine learning model to categorize the readiness levels of PLK students at Universitas Negeri Padang. K-Means clustering was initially applied to group evaluation data into three distinct readiness categories, with the resulting cluster labels serving as target variables. Subsequently, the C4.5 algorithm was utilized to extract decision rules and generate a predictive decision tree. The dataset comprises 3,578 student grade records obtained

from the Center for Field Experience (Pusat PPL) at Universitas Negeri Padang for the July–December 2025 period, encompassing ten assessment indicator variables ranging from partner school observations to final practice reports. K-Means clustering classified 2,619 students (73.20%) into the "Ready" category, 936 students (26.16%) as "Sufficiently Ready," and 23 students (0.64%) as "Needs Mentoring". The C4.5 model selected the learning media indicator (X4) as the root node, achieving an overall accuracy of 92.32%, a weighted precision of 92.33%, a recall of 92.32%, and an F1-score of 92.33%. These findings demonstrate that the hybrid K-Means–C4.5 approach provides an objective, transparent, and data-driven instrument to support PLK administrators in final evaluations and in determining student readiness status upon completing field practice.

PENDAHULUAN

Praktik Lapangan Kependidikan (PLK) merupakan wahana strategis bagi mahasiswa program studi kependidikan untuk membentuk kompetensi profesional sebagai calon guru melalui pengalaman mengajar langsung di sekolah mitra (Khoiriyah et al., 2025). Keberhasilan pelaksanaan PLK terbukti berkontribusi terhadap peningkatan minat mahasiswa untuk berprofesi sebagai guru, sehingga kualitas pembimbingan dan penilaian selama praktik menjadi faktor penentu keberhasilan program ini (Nopel & Afroo, 2025). Meskipun demikian, pelaksanaan PLK di lapangan masih dihadapkan pada berbagai kendala teknis dan administratif yang berpotensi memengaruhi objektivitas penilaian akhir mahasiswa (Afrianti & Mariya, 2024).

Kesenjangan kompetensi antarmahasiswa dalam menjalankan tugas kependidikan di lapangan, misalnya dalam pengelolaan kelas, menunjukkan bahwa penilaian kinerja PLK perlu memperhitungkan banyak indikator secara simultan agar status akhir yang diinput benar-benar mencerminkan kondisi riil mahasiswa (Nasution & Rivauzi, 2025). Pada praktiknya, sistem penilaian PLK di Universitas Negeri Padang merekap nilai akhir mahasiswa dari dosen pembimbing dan guru pamong setelah mahasiswa menyelesaikan seluruh rangkaian praktik di sekolah mitra, namun proses rekapitulasi dan penetapan status kesiapan tersebut masih dilakukan secara manual tanpa dukungan analisis data yang terstandar (PPL UNP, 2023). Padahal, pemanfaatan analitik data terbukti mampu meningkatkan objektivitas dan konsistensi evaluasi pendidikan dibandingkan penilaian manual yang rentan terhadap bias subjektif penilai (Ahmed et al., 2025).

Perkembangan Machine Learning turut mendorong pemanfaatan teknik klasterisasi dan klasifikasi untuk mengubah data mentah menjadi pengetahuan yang dapat mendukung pengambilan keputusan. K-Means Clustering merupakan salah satu algoritma unsupervised learning yang paling banyak digunakan karena mampu mengelompokkan objek data ke dalam sejumlah klaster berdasarkan kedekatan jarak antarobjek terhadap titik pusat klaster (centroid), sehingga cocok digunakan untuk mengelompokkan data penilaian tanpa memerlukan label kelas di awal (Annas & Wahab, 2023). Adapun C4.5 merupakan algoritma supervised learning berbasis pohon keputusan yang bekerja dengan menghitung nilai Gain Ratio pada setiap atribut untuk menentukan simpul akar dan menyusun aturan klasifikasi yang mudah diinterpretasikan, sehingga banyak dimanfaatkan pada kasus prediksi maupun penentuan kelulusan berdasarkan data akademik (Nuari Anisa Sivi et al., 2023). Kemampuan algoritma klasifikasi berbasis pohon keputusan dalam mengenali pola kesalahan maupun kecenderungan capaian peserta didik turut menjadikannya relevan untuk berbagai kasus analisis data Pendidikan (Diningrat et al., 2025).

Pendekatan hybrid yang menggabungkan algoritma pengklusteran dan klasifikasi telah banyak diterapkan untuk mengatasi keterbatasan masing-masing metode apabila digunakan secara terpisah. Dalam penelitian ini, hasil pengelompokan dari algoritma K-Means difungsikan sebagai label target bagi algoritma C4.5, sehingga proses klasifikasi tidak lagi bergantung pada pelabelan manual yang rentan terhadap bias subjektivitas. Sebagai gantinya, penentuan label didasarkan pada pola pengelompokan objektif yang terbentuk secara alami dari struktur data itu sendiri. Pendekatan

hibrida ini sangat relevan diterapkan pada data penilaian Praktik Lapangan Kependidikan (PLK) yang belum memiliki label status kesiapan. Melalui mekanisme ini, K-Means secara otomatis membentuk label kluster yang objektif berbasis data, sebelum diproses lebih lanjut oleh C4.5 menjadi rangkaian aturan keputusan yang transparan, terukur, dan mudah diinterpretasikan.

Penerapan model hybrid K-Means dan C4.5 telah banyak dilakukan pada berbagai domain di luar bidang pendidikan. Penelitian dengan kombinasi K-Means dan C4.5 untuk membangun model prediksi risiko pada operasi keuangan perusahaan dan menunjukkan bahwa integrasi kedua algoritma tersebut mampu meningkatkan akurasi prediksi dibandingkan model Tunggal (Pan, 2024). Pada ranah pemasaran, mengkombinasikan K-Means untuk segmentasi konsumen dan C4.5 untuk mengklasifikasikan minat belanja konsumen melalui iklan media sosial (Banjarnahor et al., 2024), perbandingan performa K-Means dan C4.5 pada data penjualan dan menemukan bahwa pendekatan hybrid menghasilkan pengelompokan yang lebih representative (Wijaya et al., 2023). Penerapan serupa juga dilaporkan menggunakan K-Means dan C4.5 untuk menganalisis kluster dampak kesehatan akibat pandemi Covid-19 (Pangan et al., 2025).

Pada bidang pendidikan, penerapan hybrid K-Means dan C4.5 juga mulai berkembang, misalnya pada prediksi penerima beasiswa mahasiswa (Wandri et al., 2023), segmentasi performa mahasiswa pada mata kuliah tertentu menggunakan K-Means Clustering (Maharani et al., 2024), deteksi dini mahasiswa berisiko putus studi melalui kombinasi K-Means dan decision tree (Akbar et al., 2025), serta peningkatan akurasi klasterisasi capaian akademik menggunakan algoritma pendukung tambahan (Darmayanti et al., 2024). Kajian tinjauan yang dilakukan turut menegaskan bahwa pendekatan Machine Learning, termasuk kombinasi klasterisasi dan klasifikasi, semakin banyak dimanfaatkan untuk menganalisis data pendidikan guna mendukung keputusan yang lebih objektif dibandingkan evaluasi manual (Ersozlu et al., 2024). Namun demikian, seluruh penelitian tersebut masih berfokus pada domain keuangan, pemasaran, kesehatan, maupun prediksi akademik umum seperti kelulusan dan penerimaan beasiswa, sehingga penerapannya pada konteks penilaian kinerja praktik keguruan belum banyak dikaji.

Model hybrid K-Means dan C4.5 telah terbukti efektif pada berbagai domain, namun penerapannya dalam mengevaluasi kinerja praktik keguruan masih sangat terbatas. Secara khusus, belum ada studi yang menerapkan pendekatan sekuensial ini untuk mengklasifikasikan status kesiapan mahasiswa Praktik Lapangan Kependidikan (PLK) berbasis rekapitulasi nilai akhir. Rekapitulasi penilaian dari dosen pembimbing dan guru pamong pascapraktik selama ini masih berisiko memunculkan bias evaluasi, sehingga memerlukan mekanisme klasifikasi yang objektif dan terukur. Kesenjangan (gap) tersebut mendasari kebaruan (novelty) utama penelitian ini, yaitu mengadaptasi algoritma hybrid ke dalam konteks evaluasi PLK yang memiliki karakteristik indikator kompetensi pedagogis tersendiri. Penelitian ini bertujuan membangun model hybrid K-Means dan C4.5 untuk mengklasifikasikan status kesiapan mahasiswa PLK di Universitas Negeri Padang berdasarkan sepuluh indikator kompetensi akhir, sekaligus menentukan atribut yang paling dominan memengaruhi hasil klasifikasi tersebut. Tujuan dari penelitian ini adalah untuk menghasilkan arsitektur pemodelan yang memanfaatkan pengklusteran K-Means guna membentuk label kelas secara otomatis dan objektif, yang selanjutnya dilatih dengan C4.5 untuk menghasilkan aturan keputusan yang transparan dan akurat.

Kontribusi studi ini terletak pada pengintegrasian klusterisasi K-Means dan klasifikasi C4.5 yang mengubah data penilaian tanpa label menjadi aturan keputusan yang transparan, terukur, dan berbasis data. Temuan ini menunjukkan bahwa pendekatan hybrid K-Means-C4.5 dapat menyediakan instrumen yang objektif, transparan, dan berbasis data untuk membantu pengelola PLK dalam penilaian akhir PLK dan menetapkan status kesiapan mahasiswa setelah menyelesaikan seluruh rangkaian praktik di sekolah mitra.

LANDASAN TEORI

Kesiapan Praktek Lapangan Kependidikan

Praktek Lapangan Kependidikan (PLK) merupakan program wajib bagi mahasiswa studi kependidikan untuk mengaplikasikan teori pedagogis dan penguasaan materi di lingkungan sekolah secara nyata. Kesiapan mahasiswa dalam menempuh PLK mencakup kompetensi akademik, penguasaan materi bidang studi, kesiapan mental, keterampilan mengajar (pedagogik), serta pemenuhan kriteria administratif. Evaluasi kesiapan yang akurat dan berbasis data diperlukan agar pemetaan mahasiswa yang layak diterjunkan ke lapangan dapat dilakukan secara efektif (Adhia dkk., 2023).

Machine Learning

Salah satu pendekatan komputasional yang dapat diterapkan untuk efisiensi evaluasi tersebut adalah integrasi algoritma Machine Learning. Melalui teknik pembelajaran data historis, algoritma seperti K-Means Clustering dapat digunakan untuk mengelompokkan tingkat kesiapan mahasiswa secara objektif, yang kemudian dikombinasikan dengan algoritma Decision Tree (C4.5) untuk menghasilkan pola dan aturan keputusan yang terstruktur (Kadhim Gharkan & Radif, 2024).

K-Means Clustering

K-Means Clustering diterapkan pada tahap awal untuk mengelompokkan data nilai akhir mahasiswa PLK ke dalam tiga status kesiapan, yaitu Siap, Cukup Siap, dan Perlu Pendampingan, tanpa memerlukan label kelas pada data awal. Pengelompokan dilakukan dengan menghitung kedekatan jarak setiap data terhadap titik pusat kluster (centroid) menggunakan jarak Euclidean pada Persamaan (1), kemudian meminimalkan total jarak kuadrat dalam kluster (within-cluster sum of squares) pada Persamaan (2) (Annas & Wahab, 2023).

$$d(x, c) = \sqrt{\sum_{i=1}^n (x_i - c_i)^2} \quad (1)$$

$$\arg \min C \sum_{i=1}^k \sum_{x \in C_i} \|x - \mu_j\|^2 \quad (2)$$

Notasi x_i pada kedua persamaan tersebut merepresentasikan vektor nilai akhir mahasiswa ke- i , c_k adalah titik pusat kluster ke- k , dan n menyatakan jumlah indikator kompetensi yang digunakan dalam perhitungan jarak. Pada Persamaan (2), K menyatakan jumlah kluster yang ditetapkan dan J merepresentasikan total jarak kuadrat seluruh data terhadap centroid klasternya, yang nilainya diminimalkan pada setiap iterasi algoritma

Dalam proses klasterisasi, algoritma K-Means bekerja melalui serangkaian tahapan iteratif untuk menghasilkan kelompok data yang memiliki tingkat kemiripan tinggi dalam satu kluster dan perbedaan yang jelas dengan kluster lainnya. Setiap tahapan dirancang untuk memperbarui posisi pusat kluster (centroid) secara bertahap hingga diperoleh struktur kluster yang stabil. Proses iteratif tersebut bertujuan untuk meminimalkan variasi intra-kluster dan meningkatkan pemisahan antar kluster (Kaur et al., 2024). berdasarkan mekanisme dasar algoritma K-Means (Ikotun Abiodun M., 2023), tahapan umum algoritma K-Means dapat dijelaskan sebagai berikut:

1. Menentukan jumlah kluster $K=3$ berdasarkan tiga kategori status kesiapan yang telah ditetapkan pada instrumen penilaian PLK.
2. Menetapkan titik pusat kluster (centroid) awal secara acak dari data nilai akhir mahasiswa.
3. Menghitung jarak setiap data terhadap ketiga centroid menggunakan Persamaan (1).
4. Mengalokasikan setiap data ke kluster dengan jarak terdekat.

5. Memperbarui posisi centroid berdasarkan rata-rata anggota tiap kluster.
6. Mengulangi langkah 3 hingga 5 sampai posisi centroid konvergen dan tidak ada lagi perubahan anggota kluster.

Label kluster yang dihasilkan pada tahap ini merepresentasikan status kesiapan mahasiswa secara objektif. Penentuan status tersebut didasarkan pada kedekatan karakteristik nilai akhir yang diperoleh mahasiswa. Selanjutnya, label kelas yang terbentuk dijadikan sebagai target pada proses klasifikasi menggunakan algoritma C4.5.

Klasifikasi C4.5

Algoritma C4.5 diterapkan pada tahap kedua untuk membangun pohon keputusan yang menjelaskan atribut paling dominan dalam menentukan status kesiapan mahasiswa berdasarkan label kluster yang dihasilkan K-Means. Proses ini diawali dengan menghitung nilai entropy total pada Persamaan (3), dilanjutkan dengan menghitung information gain setiap atribut pada Persamaan (4), split info pada Persamaan (5), dan gain ratio pada Persamaan (6) sebagai dasar pemilihan simpul akar (Wandri et al., 2023).

$$Entropy(S) = \sum_{i=1}^n -P_i * \log_2(P_i) \quad (3)$$

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy(S_i) \quad (4)$$

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy(S_i) \quad (5)$$

$$Gain Ratio(S, A) = \frac{Gain(S, A)}{Split Information(S, A)} \quad (6)$$

Pada keempat persamaan tersebut, S menyatakan himpunan data pada suatu simpul dan A merupakan atribut kompetensi yang diuji. Notasi pi pada Persamaan (3) adalah proporsi data kelas ke-i terhadap total data pada S, sedangkan Sv dan Si pada Persamaan (4) dan (5) masing-masing merepresentasikan partisi data hasil pembagian S berdasarkan nilai v atau i dari atribut A. Gain Ratio pada Persamaan (6) diperoleh dari rasio antara Information Gain dan Split Info, sehingga atribut dengan nilai unik yang terlalu banyak tidak secara otomatis mendominasi pemilihan simpul akar. Tahapan algoritma C4.5 pada penelitian ini dilakukan melalui langkah-langkah berikut:

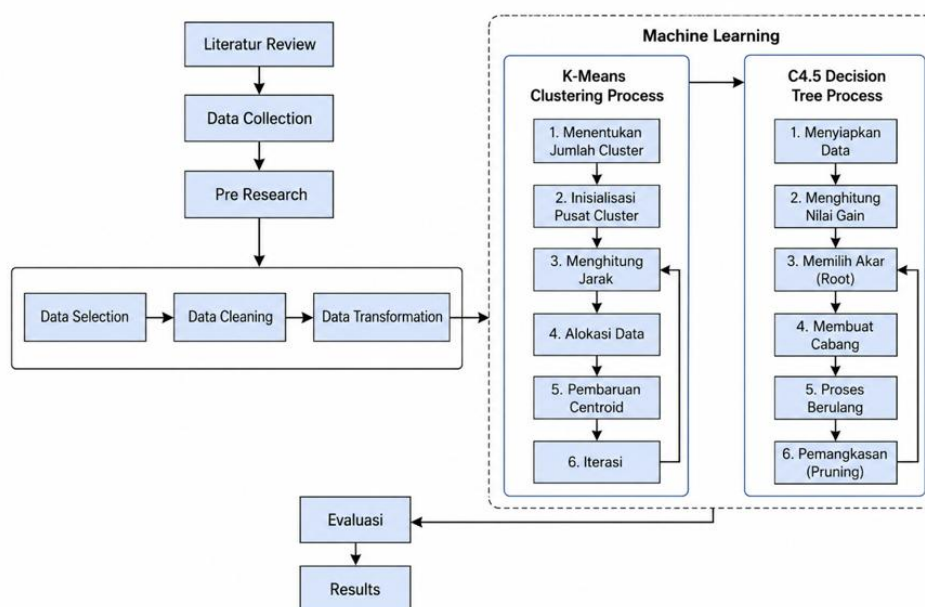
1. Menghitung nilai entropy total berdasarkan distribusi label status kesiapan menggunakan Persamaan (3).
2. Menghitung nilai information gain setiap atribut (X1 sampai X10) menggunakan Persamaan (4).
3. Menghitung nilai split info setiap atribut menggunakan Persamaan (5).
4. Menghitung nilai gain ratio setiap atribut menggunakan Persamaan (6).
5. Memilih atribut dengan nilai gain ratio tertinggi sebagai simpul akar (root node).
6. Mengulangi langkah 1 hingga 5 pada setiap cabang hingga diperoleh simpul akhir (leaf node) yang murni atau kriteria penghentian terpenuhi.

Hasil akhir dari tahap ini berupa struktur pohon keputusan dan aturan IF-THEN yang menjelaskan kombinasi atribut penentu status kesiapan mahasiswa. Aturan tersebut menyajikan pola keterkaitan antar-indikator secara transparan dan mudah diinterpretasikan. Dengan demikian, hasil analisis ini dapat digunakan sebagai dasar objektif dalam menginput nilai akhir sekaligus menetapkan status kesiapan mahasiswa PLK.

METODE PENELITIAN

Berdasarkan permasalahan dan tujuan yang telah diuraikan pada bagian pendahuluan, penelitian ini membangun model hybrid K-Means dan C4.5 untuk mengklasifikasikan status kesiapan mahasiswa PLK di Universitas Negeri Padang berdasarkan sepuluh indikator kompetensi pada nilai akhir praktik, kemudian mengekstraksi aturan keputusan yang menjelaskan atribut paling dominan dalam menentukan status tersebut, sehingga hasilnya dapat langsung dimanfaatkan sebagai dasar objektif dalam menginput nilai akhir dan status kesiapan mahasiswa setelah PLK selesai dilaksanakan.

Penelitian ini menggunakan pendekatan kuantitatif dengan desain eksperimen komputasional, karena seluruh proses analisis dilakukan terhadap data numerik hasil penilaian PLK melalui perhitungan matematis dan algoritmik, bukan melalui interpretasi naratif atas fenomena sosial. Ranah penelitian ini berada pada domain penerapan Machine Learning untuk evaluasi pendidikan (educational data mining), dengan cakupan dibatasi pada 3.578 rekaman nilai akhir mahasiswa PLK Universitas Negeri Padang periode Juli-Desember 2025 beserta sepuluh variabel indikator kompetensi (X1 sampai X10) yang telah ditetapkan pada instrumen penilaian PLK. Adapun gambaran metodologi dapat disajikan dalam kerangka penelitian pada gambar 1.



Gambar 1. Kerangka Penelitian

Kerangka penelitian pada Gambar 1 menunjukkan bahwa proses dimulai dari studi literatur, pengumpulan data nilai PLK, dan dilanjutkan dengan pre-research melalui tahapan data selection, data cleaning, dan data transformation. Data yang telah bersih dan terstandarisasi kemudian diproses menggunakan algoritma K-Means untuk membentuk label status kesiapan, yang selanjutnya dijadikan target pada algoritma C4.5 untuk membangun pohon keputusan. Model yang dihasilkan diuji menggunakan confusion matrix sebelum hasil klasifikasinya digunakan untuk menginput nilai akhir dan status kesiapan mahasiswa.

HASIL DAN PEMBAHASAN

Hasil Penorepan Hybrid K-Means dan C4.5

Tahapan analisis data dalam bagian ini dijalankan secara sistematis mengikuti kerangka penelitian yang disajikan pada Gambar 1. Proses pengolahan diawali dengan pemetaan karakteristik dataset, pengelompokan berbasis klasterisasi K-Means, dilanjutkan klasifikasi pohon

keputusan C4.5, serta diakhiri dengan pengujian kinerja model menggunakan confusion matrix. Seluruh uraian hasil dari tiap-tiap tahapan tersebut dipaparkan secara runtut melalui point 3.1 hingga 3.5. berikut.

Dataset Penelitian

Dataset penelitian ini bersumber dari rekap nilai akhir PLK yang dikelola Pusat PPL Universitas Negeri Padang periode Juli-Desember 2025, dengan dimensi awal 3.578 baris data dan sepuluh variabel indikator kompetensi (X1 sampai X10) setelah tahap data selection mereduksi atribut identitas mahasiswa. Sebagian nilai pada variabel observasi dan laporan akhir masih berupa teks "BL" (Belum Lengkap) akibat komponen penilaian yang belum dilengkapi oleh dosen pembimbing atau guru pamong pada saat data diekstraksi, sehingga nilai tersebut memerlukan penanganan lebih lanjut pada tahap data cleaning sebelum dapat diproses oleh algoritma klasterisasi. Adapun sampel data disajikan pada Tabel 1.

Tabel 1. Sampel Data Nilai Praktek Lapangan Kependidikan (PLK)

Mhs	Observasi	LMT	RMP	Media	LKPD	LMM	Manaj	Ekstrakurikuler	PTK	Laporan
1	96	98,4	96	96	95	95,6	95	90	96	87,5
2	94	89,8	80	80	87,5	81,6	90,63	93,128	86,5	96,88
3	BL	100	100	100	100	97	100	100	72,5	90,63
4	BL	100	98	100	100	100	100	100	99	BL
5	90	90	90	90	90	94	90,63	90,63	92,5	87,5
6	90	75	90	85	87,5	90	93,75	96,88	80	75
7	95	79	95	95	95	94	93,75	96,88	94,5	96,88
8	85	74,4	85,5	90	83	68,6	58,128	80,628	84	93,75
9	85	86,6	90	89	90	92	90,63	88,75	90	90,63
10	90	69	92,5	75	95	93	87,5	87,5	72,5	84,38

Sebagaimana terlihat pada Tabel 1, sebagian besar nilai berada pada rentang skor tinggi antara 68 hingga 100, yang mencerminkan capaian kompetensi mahasiswa pada berbagai indikator praktik mengajar di sekolah mitra. Anomali nilai "BL" pada Tabel 1 dominan tersebar pada variabel observasi (X1) dan laporan akhir (X10) akibat asinkronisasi input nilai oleh dosen pembimbing lapangan atau guru pamong pada tenggat waktu penarikan data dari sistem PPL UNP. Dengan mempertahankan struktur matriks utuh berukuran 3.578 baris dan 10 kolom, dataset ini siap digunakan untuk melihat sebaran kecenderungan nilai akhir performa mengajar mahasiswa secara kelembagaan sebelum diproses lebih lanjut pada tahap klasterisasi.

Analisis Variabel

Proses evaluasi performa dan kesiapan mahasiswa dalam pelaksanaan Praktik Lapangan Kependidikan (PLK) diukur berdasarkan 10 indikator penilaian utama yang mempresentasikan kompetensi praktikan di lapangan, yaitu: Observasi Sekolah Mitra (X1), Latihan Mengajar Terbimbing (X2), Rencana Pembelajaran dan Modul Ajar (X3), Media Pembelajaran (X4), Lembar Kegiatan Peserta Didik/LKPD (X5), Latihan Mengajar Mandiri (X6), Kemampuan Manajemen Persekolahan (X7), Kegiatan Non-Pembelajaran (X8), Penelitian Tindakan Kelas/Mini Riset (X9), dan Laporan Akhir PLK (X10). Adapun sampel data nilai dari kesepuluh indikator tersebut disajikan pada Tabel 2.

Tabel 2. Sample Data Nilai Indikator Penilaian Praktek Lapangan Kependidikan (PLK)

X1	X2	X3	X4	X5	X6	X7	X8	X9	X10
96	98,4	96	96	95	95,6	95	90	96	87,5
94	89,8	80	80	87,5	81,6	90,63	93,128	86,5	96,88
BL	100	100	100	100	97	100	100	72,5	90,63
BL	100	98	100	100	100	100	100	99	BL
90	90	90	90	90	94	90,63	90,63	92,5	87,5
90	75	90	85	87,5	90	93,75	96,88	80	75
95	79	95	95	95	94	93,75	96,88	94,5	96,88
85	74,4	85,5	90	83	68,6	58,128	80,628	84	93,75
85	86,6	90	89	90	92	90,63	88,75	90	90,63
90	69	92,5	75	95	93	87,5	87,5	72,5	84,38

Seluruh indikator X1 hingga X10 pada Tabel 2 saling melengkapi dalam mengukur tingkat ketercapaian aspek akademis, praktis, serta manajerial mahasiswa selama menjalani seluruh tahapan kegiatan Praktik Lapangan Kependidikan di sekolah mitra. Variasi nilai yang tercatat pada setiap atribut menjadi variabel input utama yang diolah dalam tahapan analisis data berikutnya.

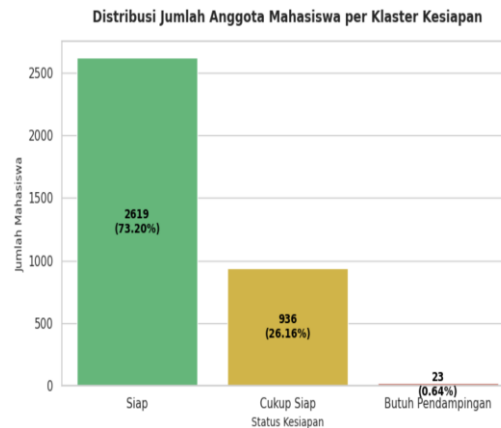
Hasil Metode 1 : Klasterisasi K-Means

K-Means Clustering diterapkan terhadap 3.578 data nilai akhir PLK dengan menghitung jarak Euclidean dari setiap data mahasiswa terhadap tiga titik pusat klaster (centroid), yaitu Jarak_C0, Jarak_C1, dan Jarak_C2, sebagaimana diuraikan pada Subbab 2.1. Setiap data kemudian dialokasikan ke klaster dengan nilai jarak centroid paling minimum/terdekat, sehingga setiap mahasiswa memperoleh label klaster (0, 1, atau 2) serta status kesiapan yang sesuai. Sampel hasil perhitungan jarak Euclidean dan penetapan klaster disajikan pada Tabel 3.

Tabel 3. Perhitungan jarak Euclidean dan Penentuan Cluster K-Means

X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	JarakC0	JarakC1	JarakC2	Klaster	Status_Kesiapan
96	98,4	96	96	95	95,6	95	90	96	87,5	10,69	28,42	277,56	0	Siap
94	89,8	80	80	87,5	81,6	90,63	93,128	86,5	96,88	29,66	18,44	257,29	1	Cukup Siap
0	100	100	100	100	97	100	100	72,5	90,63	95,53	91,59	279,32	1	Cukup Siap
0	100	98	100	100	100	100	100	99	0	130,8	127,6	273,07	1	Cukup Siap
90	90	90	90	90	94	90,63	90,63	92,5	87,5	13,80	14,11	265,76	0	Siap
90	75	90	85	87,5	90	93,75	96,88	80	75	32,76	22,51	254,31	1	Cukup Siap
95	79	95	95	95	94	93,75	96,88	94,5	96,88	15,35	25,32	276,80	0	Siap
85	74,4	85,5	90	83	68,6	58,128	80,628	84	93,75	56,53	37,67	235,89	1	Cukup Siap
85	86,6	90	89	90	92	90,63	88,75	90	90,63	17,81	8,98	262,68	1	Cukup Siap
90	69	92,5	75	95	93	87,5	87,5	72,5	84,38	41,31	27,71	249,92	1	Cukup Siap

Berdasarkan hasil pengelompokan secara keseluruhan terhadap 3.578 data mahasiswa, terbentuk distribusi anggota per klaster kesiapan yang divisualisasikan pada Gambar 2. Sebagian besar mahasiswa berada pada kategori Siap dengan proporsi mencapai 73,20% atau sebanyak 2.619 orang. Sementara itu, kategori Cukup Siap mencakup 936 mahasiswa (26,16%) dan kategori Butuh Pendampingan hanya terdiri dari 23 mahasiswa (0,64%)



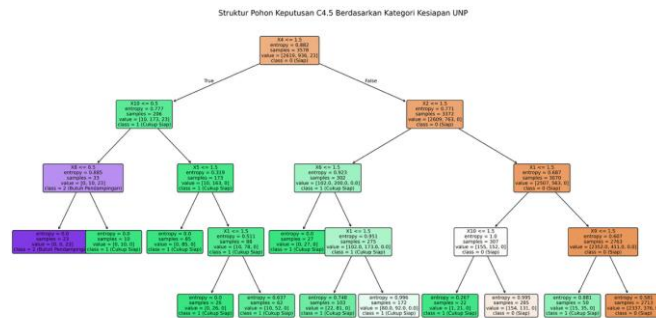
Gambar 2. Distribusi Mahasiswa perCluster

Pola sebaran ini mengindikasikan bahwa sebagian besar mahasiswa PLK Universitas Negeri Padang memiliki tingkat kompetensi mengajar yang memadai dan berada pada zona nilai tinggi (mendekati centroid 0). Sementara itu, label klaster yang terbentuk pada tahap ini selanjutnya diwariskan sebagai target kelas (ground truth) pada proses pengujian dan klasifikasi menggunakan C4.5 yang dibahas pada subbab berikutnya.

Hasil Metode 2

Label klaster hasil K-Means selanjutnya dijadikan target kelas pada algoritma C4.5. Perhitungan Gain Ratio pada Node 0 menunjukkan bahwa atribut media pembelajaran (X4) memperoleh nilai gain tertinggi dan terpilih sebagai simpul akar (root node), meskipun atribut lembar kegiatan peserta didik (X5) mencatatkan nilai Gain Ratio tertinggi pada perhitungan awal. Pemilihan X4 sebagai root node terbukti mampu memisahkan seluruh mahasiswa kategori Perlu Pendampingan hanya dalam satu kali percabangan, sehingga menghasilkan struktur pohon keputusan yang lebih ringkas dan mudah diinterpretasikan dibandingkan bila X5 dipaksakan sebagai pemisah utama.

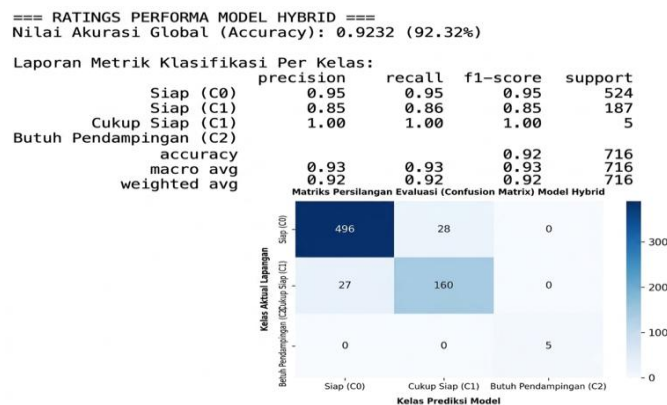
Struktur pohon pada Gambar 3 lebih dominan mengarah pada percabangan kelas Siap, terutama pada cabang kanan pohon ($X4 > 1,5$) yang berujung pada simpul akhir bervolume 2.713 sampel dengan mayoritas kelas Siap, atau setara 76 persen dari total data. Sebaliknya, percabangan menuju kelas Perlu Pendampingan hanya terbentuk pada satu simpul akhir bervolume 23 sampel di ujung cabang kiri pohon, yang menegaskan bahwa mayoritas mahasiswa PLK berada pada kondisi kompetensi yang memadai, sementara kelompok yang benar-benar memerlukan pendampingan intensif teridentifikasi secara terpisah dan spesifik melalui kombinasi atribut X4, X10, dan X8. Keunggulan pemilihan X4 sebagai pemisah utama juga tampak dari penurunan nilai entropy populasi dari 0,882 pada simpul akar menjadi 0,777 pada cabang kiri dan 0,771 pada cabang kanan, yang menunjukkan bahwa pemisahan berbasis X4 secara efektif meningkatkan kemurnian kelas pada setiap cabang berikutnya



Gambar 3. Visualisasi Pohon Keputusan C4.5

Hasil Evaluasi

Model hybrid ini bekerja secara sekuensial dan saling melengkapi. K-Means berperan membentuk label status kesiapan secara objektif dari data nilai akhir yang tidak berlabel, menggantikan penetapan kelas secara manual. Label tersebut kemudian diwariskan kepada C4.5 untuk menghasilkan aturan keputusan eksplisit yang menjelaskan atribut penentu status kesiapan. Kombinasi ini menutupi keterbatasan kedua algoritma, di mana K-Means menyediakan target kelas dan C4.5 memberikan mekanisme klasifikasi yang mudah diinterpretasikan. Performa pengujian model pada 716 data uji disajikan pada Gambar 4.



Gambar 4. Matrix Performa Model

Berdasarkan Gambar 4, Pengujian model hybrid pada 716 data uji menghasilkan akurasi global sebesar 92,32% dengan weighted average untuk precision, recall, dan f1-score sebesar 0,92. Kategori Siap (C0) mencatatkan f1-score 0,95 (524 sampel), kategori Cukup Siap (C1) sebesar 0,85 (187 sampel), dan kategori Butuh Pendampingan (C2) meraih nilai sempurna 1,00 (5 sampel). Kesalahan klasifikasi hanya terjadi antara kelas C0 dan C1 (total 55 data), sedangkan kelas C2 terprediksi tanpa kesalahan. Hasil ini membuktikan bahwa model tidak hanya akurat secara keseluruhan, tetapi juga sangat andal dalam mengidentifikasi kelompok mahasiswa yang membutuhkan pendampingan khusus.

KESIMPULAN DAN SARAN

Kesimpulan

Berdasarkan hasil perancangan, pengujian, dan pembahasan yang telah dilakukan, maka dapat disimpulkan bahwa implementasi Machine Learning pada 3.578 data mahasiswa berhasil mengalihkan skema evaluasi dari pola subjektif menjadi sistem berbasis data (data-driven). Pengolahan sepuluh indikator kompetensi secara terstandarisasi mampu meminimalkan bias penilaian antar-evaluator, sehingga menghasilkan instrumen evaluasi yang lebih objektif dan akuntabel. Kombinasi K-Means dan C4.5 terbukti efektif memproses data secara bertahap. Tahap

unsupervised learning membagi data ke dalam kluster Siap, Cukup Siap, dan Butuh Pendampingan, yang selanjutnya dilatih menggunakan C4.5 hingga meraih akurasi 92,32%. Hal ini mengonfirmasi keandalan model dalam klasifikasi kesiapan mahasiswa. Model berhasil menyusun aturan logis (If-Then Rules) yang transparan dan mudah diinterpretasikan. Indikator Media Pembelajaran (X4) teridentifikasi sebagai simpul utama (root node), diikuti oleh variabel pendukung lainnya (X1, X2, X3, X5, X6, X7, X8, X9, X10).

Saran

Berdasarkan hasil perancangan, pengujian, dan pembahasan yang telah dilakukan, terdapat beberapa hal yang masih bisa dikembangkan lebih lanjut agar hasil penelitian ini memberi manfaat yang lebih luas. Oleh karena itu, berikut beberapa saran yang dapat dipertimbangkan untuk pemanfaatan dan pengembangan penelitian ini ke depannya. Pengelola PLK UNP disarankan memakai aturan keputusan dari model ini sebagai pedoman penetapan nilai akhir agar lebih cepat dan seragam pascapraktik. Dominasi indikator X4 dapat dimanfaatkan sebagai sinyal awal pemantauan saat rekapitulasi nilai berlangsung. Selain itu, mahasiswa berpredikat "Butuh Pendampingan" dapat terdeteksi lebih awal untuk mengikuti pembekalan remedial. Model komputasi ini sebaiknya diintegrasikan ke portal Sistem Informasi PPL pada modul rekapitulasi nilai akhir. Integrasi otomatis ini memangkas proses pengolahan data manual bagi ribuan mahasiswa, sekaligus dapat dikembangkan menjadi dashboard capaian kompetensi bagi pimpinan fakultas. Penelitian mendatang disarankan memperluas cakupan dataset dengan menggabungkan data dari beberapa periode untuk menguji konsistensi posisi X4 sebagai root node. Penambahan variabel baru, seperti karakteristik sekolah mitra dan bidang studi, juga perlu dieksplorasi guna meningkatkan daya generalisasi model.

DAFTAR PUSTAKA

- Afrianti, P., & Mariya, S. (2024). Kendala Pelaksanaan Praktik Lapangan Kependidikan (PLK) Mahasiswa Pendidikan Geografi Angkatan Tahun 2020 Universitas Negeri Padang.
- Ahmed, W., Wani, M. A., Plawiak, P., Meshoul, S., Mahmoud, A., & Hammad, M. (2025). Machine learning-based academic performance prediction with explainability for enhanced decision-making in educational institutions. *Scientific Reports*, 15(1). <https://doi.org/10.1038/s41598-025-12353-4>
- Akbar, I., Sarmita Samad, I., & Rosmiana, S. (2025). Data Mining Analysis of K-means Algorithm and Decision Tree for Early Detection of Students at Risk of Dropping Out. 7(2), 148–162. <https://doi.org/10.20895/INISTA.V7I2.1630>
- Al-Tameemi, G., Xue, J., Ali, I. H., & Ajit, S. (2024). A Hybrid Machine Learning Approach for Predicting Student Performance Using Multi-class Educational Datasets. *Procedia Computer Science*, 238, 888–895. <https://doi.org/10.1016/j.procs.2024.06.108>
- Annas, M., & Wahab, S. N. (2023). Data Mining Methods: K-Means Clustering Algorithms. *International Journal of Cyber and IT Service Management*, 3(1), 40–47. <https://doi.org/10.34306/ijcitsm.v3i1.122>
- Banjarnahor, J., Hutagalung, J. P., & Sitorus, F. J. W. (2024). Analyzing Consumer Shopping Interest via Social Media Ads with K-Means and C4.5 Algorithm. *Jurnal Sisfokom (Sistem Informasi Dan Komputer)*, 13(3), 416–421. <https://doi.org/10.32736/sisfokom.v13i3.2228>
- Darmayanti, I., Adhimah, L. F., Sadewo, R., Hidayati, N., & Subarkah, P. (2024). Improving K-Means Clustering Accuracy for Academic Success Investigation With Extreme Gradient Boosting Algorithm. *Indonesian Journal of Artificial Intelligence and Data Mining*, 7(1), 104. <https://doi.org/10.24014/ijaidm.v7i1.26657>
- Diningrat, M. S. M., Mabruroh, F., Widya, H., & Nurhamidah, N. (2025). The Application of Machine Learning Algorithms in Analyzing Students' Conceptual Error Patterns in Science Learning. *Jurnal Pendidikan Dan Ilmu Fisika*, 5(1), 54–65. <https://doi.org/10.52434/jpif.v5i1.42586>

- Ersozlu, Z., Taheri, S., & Koch, I. (2024). A review of machine learning methods used for educational data. *Education and Information Technologies*, 29(16), 22125–22145. <https://doi.org/10.1007/s10639-024-12704-0>
- Ikotun Abiodun M. (2023). K-means clustering algorithms: A comprehensive review, variants analysis, and advances in the era of big data. *Information Sciences*, 622, 178–210.
- Kaur, A., Kumar, Y., & Sidhu, J. (2024). Exploring meta-heuristics for partitional clustering: methods, metrics, datasets, and challenges. *Artificial Intelligence Review*, 57(10). <https://doi.org/10.1007/s10462-024-10920-1>
- Khoiriyah, N., Arisanti, K., & Inzah, M. (2025). Peran Praktik Pengalaman Lapangan Kependidikan Terhadap Kesiapan Menjadi Guru Pada Mahasiswa Pendidikan Agama Islam. *Edukasiana: Jurnal Inovasi Pendidikan*, 4(3), 550–558. <https://doi.org/10.56916/ejip.v4i3.1384>
- Maharani, L. P. K., Angin, A. N. P., Wardhana, P. R., & Achmad, F. (2024). Evaluasi Performa Mahasiswa pada Pembelajaran Mata Kuliah Data Analitik Menggunakan K-Means Clustering. *Jurnal Konatif: Jurnal Ilmiah Pendidikan*, 1(2). <https://doi.org/10.62203/jkijip.v1i2.43>
- Nasution, A. R., & Rivauzi, A. (2025). Faktor Penyebab Ketidakmampuan Mahasiswa dalam Pengelolaan Kelas pada Mahasiswa PLK PAI UNP Periode Juli–Desember 2024. *YASIN*, 5(5), 4321–4330. <https://doi.org/10.58578/yasin.v5i5.6787>
- Nopel, N., & Afroo, F. A. (2025). Pengaruh Praktek Lapangan Kependidikan terhadap Minat Menjadi Guru pada Mahasiswa Program Studi Pendidikan Agama Islam Universitas Negeri Padang. *YASIN*, 5(5), 5444–5456. <https://doi.org/10.58578/yasin.v5i5.7691>
- Nuari Anisa Sivi, Rudi Hartono, & Putra Hanafi. (2023). Penerapan Algoritma C4.5 untuk Prediksi Kelulusan Mahasiswa berdasarkan Data Akademik. *Polygon : Jurnal Ilmu Komputer Dan Ilmu Pengetahuan Alam*, 1(5), 01–17. <https://doi.org/10.62383/polygon.v1i5.855>
- Pan, C. (2024). Construction of Risk Prediction Models for Enterprise Finance Sharing Operations Using K-Means and C4.5 Algorithms. *International Journal of Computational Intelligence Systems*, 17(1). <https://doi.org/10.1007/s44196-024-00608-3>
- Pangan, J., Anindita, K., Agung Ngurah Bagus Adhitya Wirakusuma, A., May Pamudji, K., Wurnaning Budi, D., Komang Reka Yudha Putra, I., & History, A. (2025). Analisis Klaster Dampak Kesehatan Akibat Pandemi Covid-19 Menggunakan Metode K-Means dan C4.5 Cluster Analysis of Health Impacts Due to the Covid-19 Pandemic Using the K-Means and C4.5 Methods. <https://doi.org/10.30812/nutriology.v6i2>
- PPL UNP. (2023). Penilaian Mahasiswa PLK. https://Simpel.Unp.Ac.Id/App/Pen_penilaiandpl.