

Teacher Performance Evaluation Analysis Using K-Means Clustering Algorithm and Random Forest Classification

Dito Jurinaldo¹⁾, Musli Yanto²⁾, dan Syafri Arlis³⁾

^{1,2,3}Magister Teknik Informatika, Universitas Putra Indonesia YPTK Padang
^{1,2,3}Jl. Raya Lubuk Begalung, Lubuk Begalung Nan XX, Kecamatan Lubuk Begalung Kota Padang Sumbar
E-mail: ditojurinaldo@gmail.com¹⁾, musli_yanto@upiyptk.ac.id²⁾, syafri_arlis@upiyptk.ac.id³⁾

ABSTRACT

Teacher performance assessment is a primary parameter in determining the quality of educational institutions. Evaluation systems in many elementary schools still rely on descriptive qualitative approaches. Abundant school administrative data often remain as unprocessed archival records without further analytical utilization. This condition results in school management decision-making that lacks a strong empirical foundation. This study applies data mining technology to transform administrative data into strategic information. The research focuses on SD Negeri 12 Padang Besi and involves all active teaching staff during the current academic year. The research dataset is entirely derived from internal school records. This study excludes the use of questionnaire instruments, and in-depth interview methods are not employed in the data collection process. The analysis is strictly limited to administrative aspects, without including assessments of in-class pedagogical competence. The technical implementation utilizes the K-Means Clustering algorithm to automatically identify patterns in teacher performance grouping. This process is followed by the application of the Random Forest algorithm to measure classification accuracy based on the available administrative features. The combination of these methods produces a performance mapping that is free from human subjectivity. The analytical results provide clear performance labels for each individual teacher. This study contributes to the development of a data-driven digital evaluation model. School management can use the outputs of this research as a basis for reward allocation or targeted professional development programs. This approach ensures transparency in human resource governance within the educational environment.

Keywords: *Teacher Performance, Data Mining, K-Means Clustering, Random Forest, School Administrative Data, Educational Evaluation*

Analisis Evaluasi Kinerja Guru dengan Menggunakan Algoritma K-Means Clustering dan Klasifikasi Random Forest

ABSTRAK

Penilaian kinerja guru merupakan parameter utama dalam penentuan kualitas institusi pendidikan. Sistem evaluasi pada banyak sekolah dasar saat ini masih mengandalkan pendekatan deskriptif kualitatif. Data administrasi sekolah yang melimpah sering kali hanya berakhir sebagai tumpukan arsip tanpa pengolahan lebih lanjut. Fenomena ini menyebabkan pengambilan keputusan manajemen sekolah kurang memiliki landasan empiris yang kuat. Penelitian ini menerapkan teknologi data mining untuk mentransformasi data administrasi menjadi informasi strategis. Fokus penelitian berada pada lingkungan SD Negeri 12 Padang Besi dengan melibatkan seluruh tenaga pendidik aktif pada periode tahun ajaran berjalan. Dataset penelitian bersumber sepenuhnya dari catatan internal sekolah. Penelitian ini meniadakan penggunaan instrumen kuesioner. Metode wawancara mendalam tidak diterapkan dalam pengumpulan data. Pembatasan dilakukan pada aspek administratif murni. Penilaian kompetensi pedagogik di dalam kelas tidak menjadi bagian dari analisis ini. Implementasi teknis menggunakan algoritma K-Means Clustering untuk mengidentifikasi pola pengelompokan kinerja guru secara otomatis. Proses ini dilanjutkan dengan penerapan algoritma Random Forest untuk mengukur tingkat akurasi klasifikasi berdasarkan fitur-fitur administrasi yang tersedia. Kombinasi kedua metode ini menghasilkan pemetaan kinerja yang bebas dari subjektivitas manusia. Hasil analisis memberikan label performa yang jelas bagi setiap individu tenaga pendidik. Penelitian ini memberikan kontribusi pada model evaluasi berbasis data digital. Manajemen sekolah dapat menggunakan luaran penelitian ini sebagai dasar pemberian penghargaan atau program pembinaan khusus. Pendekatan ini memastikan transparansi dalam tata kelola sumber daya manusia di lingkungan pendidikan.

Kata Kunci: Kinerja Guru, Data Mining, K-Means Clustering, Random Forest, Data Administrasi Sekolah, Evaluasi Pendidikan

1. PENDAHULUAN

Pendidikan di Indonesia memiliki masalah utama yang dihadapi seperti kualitas proses pembelajaran di kelas yang kurang optimal (Suncaka, 2023). Kegiatan belajar mengajar terlalu didominasi oleh teori yang bersifat kognitif dan hanya bertujuan agar siswa memahami materi, namun kurang relevan dengan penerapan praktis dalam kehidupan sehari-hari. Peran guru sangat penting untuk memaksimalkan potensi dan kreativitas siswa. Selain itu, teknologi pendidikan dan media pembelajaran saat ini memiliki potensi besar sebagai alat bantu yang efektif dari jenjang pendidikan dasar hingga atas. Media pembelajaran dapat menjadikan proses belajar lebih menarik, menyenangkan, dan memungkinkan siswa untuk mengakses wawasan serta informasi yang lebih lengkap di luar batas ruang kelas tradisional (Selatan dkk., 2021)

Implementasi teknologi dalam pendidikan dinilai sangat efektif karena dapat meningkatkan daya guna, kecepatan, dan keterlibatan siswa, sekaligus menciptakan metode pembelajaran yang relevan dan sesuai dengan perkembangan zaman. Meskipun demikian, upaya memasukkan teknologi ke dalam sistem belajar di Sekolah Dasar (SD) sering terhambat oleh berbagai kendala. Tantangan utama yang menghalangi optimalisasi ini meliputi keterbatasan sarana dan prasarana (infrastruktur), rendahnya kemampuan atau literasi digital guru, dan kurangnya kesempatan bagi pendidik untuk mendapatkan pelatihan teknis (Anwar dkk., 2025).

Data mining merupakan proses analitis yang sistematis, memanfaatkan perangkat lunak komputasi untuk mengekstrak dan mengidentifikasi pola atau aturan yang signifikan dari himpunan data bervolume besar, dengan tujuan menghasilkan pengetahuan yang kredibel guna mendukung pengambilan keputusan (Suliman, 2021). Secara praktis, penambangan data (Data Mining) mengintegrasikan serangkaian teknik komputasi dan statistik mutakhir. Implementasinya melibatkan penggunaan Kecerdasan Buatan (AI), Pembelajaran Mesin (*Machine Learning*), dan metode analisis statistik untuk mengekstrak, mengidentifikasi, dan menyaring informasi yang bernilai dari basis data yang bervolume besar (Yoliadi, 2023). Data mining merupakan proses analisis untuk menemukan pola dari data dalam jumlah besar (Han dkk., 2012; Fayyad dkk., 1996). Pemanfaatan teknik Data Mining memungkinkan institusi pendidikan mendapatkan wawasan tentang preferensi gaya belajar siswa dan sekaligus menyediakan kerangka kerja yang objektif untuk pengelompokan dan mengklasifikasikan dalam menilai kinerja para guru (Zul'Aini dkk., 2024).

Algoritma K-Means merupakan sebuah teknik pengelompokan (*clustering*) yang bersifat iteratif, berfungsi untuk mempartisi himpunan data ke dalam

sejumlah klaster (*cluster*) yang kuantitasnya (nilai K) telah ditentukan di awal. Algoritma ini memiliki keunggulan karena mudah diimplementasikan, beroperasi secara relatif cepat, sangat adaptif, dan aplikasinya sangat umum dalam praktik. Algoritma K-Means merupakan metode clustering yang широко digunakan dalam berbagai bidang (Jain, 2010). Penentuan nilai K yang optimal dapat didasarkan pada pengetahuan awal tentang struktur klaster aktual dalam data, pengalaman praktisi, atau melalui eksplorasi empiris dengan mencoba beberapa nilai K (Riani dkk., 2023).

Penelitian yang dilakukan oleh Filda Febrinita dkk., (2023) berfokus pada klusterisasi hasil belajar matematika mahasiswa Teknik Informatika di salah satu universitas. Studi ini dilakukan karena adanya fakta bahwa pemilihan peminatan sering dilakukan tanpa mempertimbangkan kemampuan dasar matematis mahasiswa, padahal hal tersebut sangat penting di bidang komputer. Untuk mengatasi isu tersebut, diterapkan Algoritma *K-Means Clustering* pada 51 data mahasiswa semester empat. Atribut yang dianalisis mencakup data asal sekolah, jurusan di SMA, serta nilai mata kuliah yang berkaitan dengan matematika dasar. Hasil penelitian ini bertujuan memberikan kerangka kerja yang objektif bagi institusi dalam menentukan peminatan mahasiswa yang paling sesuai dengan potensi akademiknya.

Algoritma *Random Forest* merupakan algoritma pembelajaran mesin yang sangat banyak digunakan (prominent) dalam penyelesaian masalah klasifikasi dan regresi (Haris dkk., 2024). Integrasi algoritma Random Forest dengan data akademik yang multidimensi masih menunjukkan tingkat eksplorasi yang terbatas dalam konteks sistem pendidikan di Indonesia (Sofiyatus Zawiyah dkk., 2024). Random Forest merupakan metode ensemble yang efektif untuk klasifikasi (Breiman, 2001). Algoritma Random Forest berpotensi menjadi solusi optimal dalam mengatasi tantangan klasifikasi (Aini dkk., 2024).

Penelitian yang dilakukan oleh Mulia Mahendra Alvanof dkk., (2024) melakukan studi mengenai Penerapan Algoritma *Random Forest* dalam Deteksi dan Klasifikasi *Ransomware*. Penelitian ini sangat relevan mengingat *ransomware* adalah jenis *malware* yang menghalangi akses sistem hingga tebusan dibayar, menimbulkan ancaman serius. Metode *machine learning Random Forest* dipilih karena kelebihannya dalam menangani dataset besar, waktu pelatihan yang singkat, dan kemampuannya mengurangi risiko overfitting. Dalam prosesnya, dilakukan seleksi fitur untuk memilih 10 fitur terbaik dari 54 fitur yang tersedia sebelum model dibangun. Model *Random Forest* yang berhasil dikonstruksi menunjukkan performa yang sangat tinggi,

mampu memprediksi file *ransomware* dengan akurasi mencapai 98.79%

Berdasarkan laporan penelitian terdahulu dan kondisi *eksisting*, teridentifikasi bahwa meskipun banyak penelitian yang menggunakan data mining dalam konteks pendidikan, belum ada yang secara spesifik mengkombinasikan *K-Means Clustering* dan *Random Forest* untuk menganalisis kinerja guru berdasarkan data administrasi sekolah di tingkat sekolah dasar. Hasil penelitian diharapkan dapat memberikan model evaluasi kinerja guru yang objektif, berbasis data, dan dapat menjadi acuan untuk pengembangan profesional guru serta peningkatan kualitas pendidikan di SD Negeri 12 Padang Besi khususnya dan sekolah dasar pada umumnya.

2. RUANG LINGKUP

Cakupan permasalahan dalam penelitian ini difokuskan pada penerapan konsep data mining untuk menganalisis kinerja guru berdasarkan data administrasi sekolah di SD Negeri 12 Padang Besi. Analisis dilakukan dengan memanfaatkan algoritma *K-Means Clustering* untuk mengidentifikasi dan mengelompokkan pola kinerja guru, serta algoritma *Random Forest* untuk mengklasifikasikan tingkat kinerja guru ke dalam kategori tertentu, yaitu tinggi, sedang, dan rendah. Data yang digunakan bersumber dari dokumentasi administrasi dan rekapitulasi internal sekolah, tanpa melibatkan teknik pengumpulan data berupa kuesioner maupun wawancara mendalam. Responden penelitian dibatasi pada guru yang aktif mengajar pada tahun ajaran yang ditetapkan sebagai periode penelitian.

Penelitian ini menitikberatkan pada aspek administrasi sekolah sebagai indikator penilaian kinerja guru, sehingga tidak mencakup evaluasi terhadap proses pembelajaran di kelas, kompetensi pedagogik, maupun penilaian langsung terhadap hasil belajar peserta didik. Variabel yang dianalisis meliputi kehadiran guru (X1) serta kelengkapan perangkat pembelajaran, yang terdiri atas Rencana Pelaksanaan Pembelajaran (RPP) (X2), Silabus (X3), Program Tahunan (X4), dan Program Semester (X5). Hasil evaluasi kinerja guru dalam penelitian ini dibatasi pada keluaran model berupa klasifikasi tingkat kinerja, tanpa disertai analisis kualitatif secara mendalam. Dengan demikian, penelitian ini diharapkan dapat memberikan gambaran penerapan konsep data mining melalui algoritma *K-Means Clustering* dan *Random Forest* sebagai dasar pendukung pengambilan kebijakan serta pengembangan kinerja guru di SD Negeri 12 Padang Besi.

3. BAHAN DAN METODE

Metodologi penelitian menyajikan kerangka kerja berperan sebagai panduan sistematis yang menunjukkan tahapan-tahapan penelitian dari awal hingga akhir. Urutan langkah dalam kerangka kerja memberikangambaran alur kerja yang terstruktur agar proses penelitian dapat berjalan secara efektif, terarah,

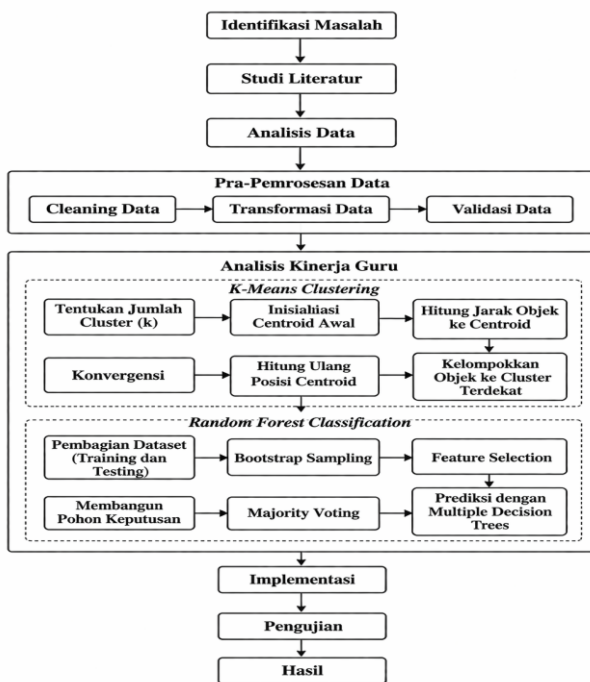
dan sesuai dengan tujuan yang telah direncanakan. Setiap tahap memiliki fungsi penting, mulai dari pengumpulan data, pengolahan, analisis, hingga penarikan kesimpulan. Kerangka kerja pada penelitian ini dapat dilihat pada Gambar 1.

Kerangka kerja penelitian ini disusun sebagai pedoman konseptual yang menampilkan alur berpikir peneliti dalam mencapai tujuan penelitian. Struktur ini menggambarkan keterkaitan antara tahapan pengolahan data, penerapan metode algoritma, hingga proses evaluasi model yang digunakan. Pada penelitian ini, kerangka kerja dirancang untuk memberikan pemahaman menyeluruh mengenai proses analisis kinerja guru berdasarkan data administrasi sekolah dengan menggunakan algoritma *K-Means Clustering* dalam tahap pengelompokan dan *Random Forest* dalam tahap klasifikasi tingkat kinerja guru. Berdasarkan Gambar 1, kerangka kerja penelitian ini dilakukan seperti melakukan Prapemrosesan data yang merupakan tahapan krusial dalam analisis data, karena kualitas hasil analisis sangat bergantung pada kebersihan, konsistensi, dan validitas data yang digunakan. Tanpa pra-pemrosesan yang baik, algoritma seperti *K-Means* dan *Random Forest* dapat menghasilkan output yang bias atau tidak akurat.

Analisis Data yang dilakukan dengan metode *K-Means Clustering* digunakan untuk mengelompokkan data berdasarkan kemiripan karakteristik, tanpa memerlukan label awal. Tahapan yang dilakukan dalam proses *K-Means* adalah pertama dengan menentukan jumlah cluster (k), selanjutnya dilakukan Inisialisasi Centroid Awal, setelah itu dilakukan Menghitung Jarak Objek ke Centroid, dan dilakukan dengan Mengecek Konvergensi, Jika konvergensi belum tercapai, posisi centroid diperbarui berdasarkan rata-rata posisi anggota cluster saat ini. Penentuan jumlah cluster juga dapat dilakukan menggunakan pendekatan seperti *elbow method* (Ariawan dkk., 2023). Proses ini memastikan bahwa pusat cluster mencerminkan distribusi data yang sebenarnya. Setelah iterasi selesai, setiap objek diklasifikasikan ke cluster terdekat. Tujuan utama dari algoritma *K-Means* adalah untuk mengidentifikasi cara optimal membagi sejumlah n entitas data ke dalam sejumlah kelompok atau cluster yang telah ditentukan. Berikut ini merupakan diagram flowchart dari algoritma *k-means* dan menggambarkan langkah-langkah dalam algoritma *k-means* dengan asumsi bahwa parameter input adalah jumlah data set sebanyak n data dan jumlah inisialisasi centroid k (Saputra & Nataliani, 2021). Pemilihan titik awal ini sangat penting karena akan memengaruhi hasil akhir dari proses clustering (Ariawan dkk., 2023). Sedangkan analisis data yang dilakukan dengan menggunakan metode *Random Forest Classification* yang berbasis ensemble learning yang sangat andal untuk klasifikasi data. Metode ini membentuk banyak pohon keputusan dan menggabungkan hasilnya untuk menghasilkan prediksi yang lebih akurat dan stabil dengan tahapan pertamanya

adalah Pembagian Dataset (Training dan Testing), selanjutnya teknik bootstrap untuk membuat subset data pelatihan secara acak, setelahnya dilakukan Feature Random Selection dimana sebagian fitur yang dipilih secara acak, dilakukan Pembentukan banyak pohon keputusan secara paralel. Hasil dari semua pohon digabungkan melalui mekanisme *majority voting*. Prediksi akhir ditentukan berdasarkan mayoritas suara dari seluruh pohon, sehingga lebih stabil dan tahan terhadap bias.

Model yang telah dibangun kemudian diterapkan pada data nyata. Implementasi ini bertujuan untuk menghasilkan rekomendasi berbasis data yang dapat digunakan dalam perencanaan, pengambilan keputusan, atau pengembangan program. Model diuji untuk mengevaluasi performa dan akurasi. Metrik evaluasi seperti akurasi, precision, recall, dan silhouette score digunakan untuk menilai kualitas hasil. Tahapan akhir adalah penyajian hasil analisis dalam bentuk visualisasi, laporan, atau rekomendasi strategis. Hasil ini tidak hanya memberikan gambaran tentang kondisi saat ini, tetapi juga membuka peluang untuk intervensi yang lebih tepat sasaran. Dengan pendekatan berbasis data, keputusan yang diambil menjadi lebih objektif dan berdampak jangka panjang



Gambar 1. Kerangka Kerja
Figure 1. Framework

Kerangka kerja penelitian ini disusun sebagai pedoman konseptual yang menampilkan alur berpikir peneliti dalam mencapai tujuan penelitian. Struktur ini menggambarkan keterkaitan antara tahapan pengolahan data, penerapan metode algoritma, hingga proses evaluasi model yang digunakan. Pada penelitian ini, kerangka kerja dirancang untuk memberikan pemahaman

menyeluruh mengenai proses analisis kinerja guru berdasarkan data administrasi sekolah dengan menggunakan algoritma K- Means Clustering dalam tahap pengelompokan dan Random Forest dalam tahap klasifikasi tingkat kinerja guru. Dapat dilihat pada gambar 1 kerangka kerja penelitian :

1. Mengidentifikasi Masalah
Tahap ini merupakan fondasi dari keseluruhan proses. Masalah yang diidentifikasi harus relevan dengan kebutuhan organisasi atau program, seperti segmentasi kelompok sasaran, klasifikasi tingkat kepuasan, atau pemetaan potensi wilayah. Pemahaman yang jelas terhadap masalah akan menentukan arah analisis dan pemilihan metode yang tepat.
2. Studi Literatur
Literatur yang dikaji memberikan landasan teoritis dan metodologis. Ini mencakup pemahaman terhadap algoritma yang digunakan, studi kasus serupa, serta pendekatan evaluatif yang telah terbukti efektif. Studi ini juga membantu menghindari kesalahan umum dan memperkuat validitas analisis.
3. Pengumpulan Data
Data dikumpulkan dari sumber yang kredibel dan relevan. Kualitas data sangat menentukan hasil analisis, sehingga proses ini harus dilakukan dengan teliti. Data primer seperti survei IKM memberikan gambaran langsung dari responden, sementara data sekunder dapat memperkaya konteks analisis.
4. Pra-Pemrosesan Data
Pra-pemrosesan data merupakan tahapan krusial dalam analisis data, karena kualitas hasil analisis sangat bergantung pada kebersihan, konsistensi, dan validitas data yang digunakan. Tanpa pra-pemrosesan yang baik, algoritma seperti K-Means dan Random Forest dapat menghasilkan output yang bias atau tidak akurat. Berikut tiga langkah utama dalam proses ini:

- 1) Cleaning Data
Langkah pertama adalah membersihkan data dari berbagai bentuk gangguan yang dapat menghambat analisis. Ini mencakup penghapusan data duplikat, penanganan nilai kosong (missing values), koreksi kesalahan input, dan identifikasi outlier. Proses ini bertujuan untuk memastikan bahwa data yang digunakan benar-benar mencerminkan kondisi nyata dan tidak mengandung noise yang dapat menyesatkan model.

- 2) Transformasi Data
Setelah data dibersihkan, langkah selanjutnya adalah mentransformasikan data ke dalam format yang sesuai dengan kebutuhan algoritma. Transformasi dapat berupa normalisasi skala numerik, encoding variabel kategorikal, atau konversi format tanggal dan

teks. Tujuannya adalah menyamakan struktur data agar dapat diproses secara efisien dan adil oleh algoritma pembelajaran mesin.

3) Validasi Data

Tahap terakhir adalah memastikan bahwa data yang telah dibersihkan dan ditransformasikan benar-benar siap untuk dianalisis. Validasi dilakukan dengan memeriksa konsistensi antar variabel, kesesuaian format, serta kelengkapan data. Proses ini juga mencakup uji logika dan deteksi anomali yang mungkin lolos dari tahap sebelumnya.

5. Analisis Data

K-Means Clustering

K-Means Clustering merupakan metode pengelompokan data yang termasuk dalam kategori unsupervised learning. Metode ini bertujuan untuk mengelompokkan objek berdasarkan tingkat kemiripan karakteristik tanpa menggunakan label kelas. Pada penelitian ini, proses *K-Means Clustering* dilakukan melalui beberapa tahapan.

1) Menentukan Jumlah Cluster (*k*)

Proses dimulai dengan menentukan jumlah cluster yang diinginkan. Penentuan nilai *k* dapat dilakukan menggunakan metode seperti *Elbow Method* atau *Silhouette Score*. Pemilihan *k* yang tepat sangat penting agar hasil pengelompokan mencerminkan struktur alami dalam data.

2) Inisialisasi *Centroid* Awal

Setelah jumlah cluster ditentukan, sistem memilih titik pusat awal (*centroid*) secara acak. Titik ini menjadi referensi awal untuk mengelompokkan data.

3) Menghitung Jarak Objek ke *Centroid*

Setiap data dihitung jaraknya ke masing-masing *centroid* menggunakan metrik seperti *Euclidean distance*. Tujuannya adalah menentukan kedekatan relatif antara data dan pusat *cluster*.

4) Mengecek Konvergensi

Proses iterasi dilakukan hingga posisi *centroid* tidak lagi berubah secara signifikan. Konvergensi menandakan bahwa struktur cluster telah stabil dan tidak perlu diperbarui lebih lanjut.

5) Menghitung Ulang Posisi *Centroid*

Jika konvergensi belum tercapai, posisi *centroid* diperbarui berdasarkan rata-rata posisi anggota cluster saat ini. Proses ini memastikan bahwa pusat *cluster* mencerminkan distribusi data yang sebenarnya.

6) Mengelompokkan Objek ke *Cluster* Terdekat

Setelah iterasi selesai, setiap objek diklasifikasikan ke cluster terdekat. Hasil akhir berupa segmentasi data yang dapat digunakan untuk menyusun strategi yang lebih tepat sasaran, seperti penyesuaian layanan, promosi, atau pengembangan program berbasis karakteristik kelompok.

Random Forest Classification

Random Forest Classification adalah algoritma pembelajaran mesin berbasis ensemble learning yang sangat andal untuk klasifikasi data. Metode ini

membentuk banyak pohon keputusan dan menggabungkan hasilnya untuk menghasilkan prediksi yang lebih akurat dan stabil. Berikut enam tahapan utamanya:

1) Pembagian Dataset (*Training* dan *Testing*)

Dataset dibagi menjadi dua bagian: data pelatihan untuk membangun model, dan data pengujian untuk mengevaluasi performa. Pembagian ini penting untuk menghindari overfitting dan memastikan model dapat bekerja baik pada data baru.

2) *Bootstrap Sampling*

Random Forest menggunakan teknik *bootstrap* untuk membuat subset data pelatihan secara acak. Setiap pohon dibangun dari sampel yang berbeda, meningkatkan keragaman dan ketahanan model terhadap noise.

3) *Feature Random Selection*

Untuk setiap pohon, hanya sebagian fitur yang dipilih secara acak. Strategi ini mencegah dominasi fitur tertentu dan meningkatkan kemampuan model dalam menangani data yang kompleks dan beragam.

4) Pembentukan *Multiple Decision Trees*

Model membentuk banyak pohon keputusan secara paralel. Setiap pohon memberikan prediksi berdasarkan subset data dan fitur yang dimilikinya, menciptakan variasi dalam hasil klasifikasi.

5) *Majority Voting*

Hasil dari semua pohon digabungkan melalui mekanisme majority voting. Prediksi akhir ditentukan berdasarkan mayoritas suara dari seluruh pohon, sehingga lebih stabil dan tahan terhadap bias.

6) Hasil Klasifikasi Akhir

Output akhir berupa label klasifikasi untuk setiap data. Hasil ini dapat digunakan untuk pengambilan keputusan berbasis data, seperti klasifikasi tingkat kepuasan, identifikasi kelompok risiko, atau prediksi keberhasilan program.

6. Implementasi

Model yang telah dibangun kemudian diterapkan pada data nyata. Implementasi ini bertujuan untuk menghasilkan rekomendasi berbasis data yang dapat digunakan dalam perencanaan, pengambilan keputusan, atau pengembangan program. Keberhasilan implementasi bergantung pada kesesuaian model dengan konteks lapangan.

7. Pengujian

Model diuji untuk mengevaluasi performa dan akurasi. Metrik evaluasi seperti akurasi, precision, recall, dan silhouette score digunakan untuk menilai kualitas hasil. Pengujian ini juga membantu mengidentifikasi kelemahan model dan memberikan dasar untuk perbaikan.

8. Hasil

Tahapan akhir adalah penyajian hasil analisis dalam bentuk visualisasi, laporan, atau rekomendasi strategis.

Hasil ini tidak hanya memberikan gambaran tentang kondisi saat ini, tetapi juga membuka peluang untuk intervensi yang lebih tepat sasaran. Dengan pendekatan berbasis data, keputusan yang diambil menjadi lebih objektif dan berdampak jangka panjang.

4. PEMBAHASAN

Berdasarkan kerangka kerja penelitian yang telah dijelaskan penelitian ini melakukan praproses data, menerapkan algoritma K-Means untuk proses clustering, menerapkan algoritma Random Forest untuk klasifikasi, dan menguji hasil menggunakan Microsoft Office Excel, Altair AI Studio, dan Visual Studio Code. Hasil penelitian ini sejalan dengan penelitian sebelumnya dalam bidang data mining pendidikan (Zul'Aini dkk., 2024).

4.1 Preprocessing

Pra-pemrosesan data merupakan tahap awal yang berfungsi untuk menyiapkan data agar dapat diolah secara optimal, memastikan kualitas. Data awal yang digunakan adalah data penilaian guru tahun 2025 yang

terdiri dari 36 entitas guru dengan 9 atribut, yaitu Kehadiran, RPP, Silabus, Program Tahunan, Program Semester, Observasi, Masa Kerja, Sertifikasi, Nilai Rata Siswa. Tahapan ini penting untuk memastikan bahwa data yang digunakan memiliki kualitas yang baik, bebas dari kesalahan, dan konsisten. Proses pra-pemrosesan mencakup tiga langkah utama, yaitu cleaning data, transformasi data, dan validasi data. Data awal guru dapat dilihat pada Tabel 1.

Berdasarkan dari Tabel 1, seluruh data nilai guru yang terdiri dari atribut Kehadiran, RPP, Silabus, Program Tahunan, Program Semester, Observasi, Masa Kerja, Sertifikasi, dan Nilai Rata Siswa dari guru. Data yang ditampilkan menunjukkan bahwa tidak terdapat nilai kosong atau anomali sehingga layak digunakan dalam tahap pemodelan selanjutnya. Dengan demikian, dataset yang telah tervalidasi ini siap untuk diolah menggunakan algoritma K-Means guna mengelompokkan nilai guru berdasarkan karakteristik akademik dan non-akademiknya. Adapun hasil cluster dapat disajikan pada Tabel 2.

Tabel 1. Data Awal Guru

Table 1. Initial Teacher Data

Teachers	Attendance	Lesson Plans	Syllabus	Annual Program	Semester Plan	Observations	Term of Service	Certification	Student Average Grade
G1	95	90	92	88	90	91	16	1	85
G2	85	80	78	82	79	81	32	1	80
G3	92	88	86	90	89	87	20	1	82
G4	80	75	70	72	74	73	17	1	80
G5	88	82	84	85	83	84	18	1	80

4.2 Analisis K-Means Clustering

K-Means Clustering adalah sebuah metode pembelajaran tanpa pengawasan (unsupervised learning) yang digunakan untuk mengelompokkan data ke dalam sejumlah cluster berdasarkan kesamaan karakteristik. Proses ini dimulai dengan menetapkan jumlah cluster yang diinginkan, di mana algoritma kemudian secara berulang (iteratif) akan mencari dan memperbarui pusat

cluster (centroid). Setiap titik data akan diukur jaraknya terhadap setiap centroid dan ditempatkan pada cluster yang terdekat. Centroid dari setiap cluster kemudian dihitung ulang (diperbarui posisinya) hingga hasil pengelompokan mencapai kondisi stabil, yaitu ketika tidak ada lagi perubahan signifikan dalam keanggotaan cluster seperti pada Tabel 2

Tabel 2. Hasil Cluster Algoritma K-Means

Table 2. K-Means Algorithm Cluster Results

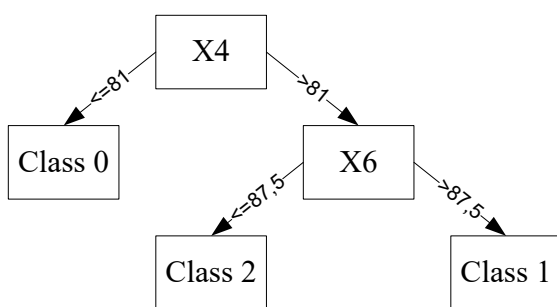
Teachers	X ₁	X ₂	X ₃	X ₄	X ₅	X ₆	X ₆	X ₇	X ₈	X ₉	C0	C1	C2	JT	Cluster
G33	89	86	85	88	87	87	6	1	92	89	31,61211	16,31291	11,54782	11,54782	2
G36	88	84	86	87	86	86	1	1	80	88	28,18514	18,59211	10,24322	10,24322	2
G5	88	82	84	85	83	84	18	1	80	88	25,88322	26,13852	8,522779	8,522779	2
G7	89	85	83	84	82	85	10	1	78	89	24,333	23,17326	5,612032	5,612032	2
G9	91	87	85	88	86	87	2	0	84	91	30,79989	14,96663	9,910055	9,910055	2

Berdasarkan Tabel 2 diatas pada iterasi ke-4, diperoleh hasil pengelompokan data dengan distribusi sebagai berikut: Cluster 0 (C0) sebanyak 13 data, Cluster 1 (C1) sebanyak 10 data, Cluster 2 (C2) sebanyak 13 data. Algoritma K-Means Clustering berhasil mengelompokkan data menjadi tiga cluster yaitu cluster 0 (C0) berisi 13 data dengan nilai guru tinggi, cluster 1

(C1) berisi 10 data dengan nilai guru sedang, dan cluster 2 (C2) berisi dengan nilai guru rendah. Sebagai tahap akhir dari proses analisis K-Means Clustering, diperoleh hasil pengelompokan data nilai guru yang menunjukkan distribusi pada masing-masing cluster. Hasil ini menggambarkan pola kesamaan karakteristik antar nilai guru berdasarkan atribut penelitian yang digunakan

4.2 Analisis Random Forest

Algoritma Random Forest termasuk kedalam pengembangan dari algoritma Decision Tree yang menggabungkan sejumlah pohon keputusan untuk meningkatkan akurasi prediksi. Setiap pohon dibangun secara independen menggunakan subset data acak yang diambil dari dataset pelatihan melalui proses bootstrap sampling. Tahapan ini digunakan untuk membangun beberapa pohon keputusan secara acak. Hasil dari masing-masing pohon keputusan kemudian digabungkan melalui proses voting. Dataset ini berasal dari 36 entri hasil dari pengelompokan data. Setiap entri memuat nilai numerik untuk variabel X_1 hingga X_9 dan output Y yang siap digunakan untuk pelatihan model.



Gambar 2. Pembentukan Pohon Pada Algoritma Random Forest

Figure 2. Tree Formation in Random Forest Algorithm

Gambar 2 menunjukkan struktur pembentukan pohon dengan memperlihatkan bagaimana setiap cabang terbentuk berdasarkan distribusi kelas pada tiap node. Struktur ini menggambarkan alur keputusan dari root hingga node-node berikutnya yang terbentuk melalui proses pemilihan atribut terbaik. Pola percabangan memberikan gambaran yang lebih intuitif mengenai proses kerja algoritma dalam mengelompokkan data

4.3 Proses Evaluasi

Evaluasi hasil prediksi digunakan untuk mengetahui tingkat performa model dalam mengklasifikasikan data berdasarkan hasil prediksi. Evaluasi ini menampilkan perbandingan antara nilai aktual dan nilai prediksi. Hasil klasifikasi menggunakan algoritma Random Forest menunjukkan bahwa model Random Forest juga menghasilkan prediksi sempurna dengan tingkat kesalahan 0%. Melihat performa model secara lebih terstruktur yang dapat dilihat pada Gambar 3.

accuracy: 100.00%

	true 0	true 1	true 2	class precision
pred. 0	3	0	0	100.00%
pred. 1	0	2	0	100.00%
pred. 2	0	0	3	100.00%
class recall	100.00%	100.00%	100.00%	

Gambar 3. Matrik Evaluasi

Figure 3. Evaluation Matrix

Gambar 3 menunjukkan bahwa baik algoritma Random Forest mampu mengklasifikasikan seluruh data dengan benar tanpa kesalahan klasifikasi. Hal ini terlihat dari nilai pada setiap kelas yang seluruhnya berada pada diagonal utama matriks, menandakan hasil prediksi sama dengan data aktual. Setiap metrik seperti akurasi, presisi, recall, dan F1-Score menunjukkan nilai 1 atau setara dengan 100%. Nilai tersebut mengindikasikan bahwa model Algoritma Random Forest memiliki performa klasifikasi yang sangat baik dan konsisten pada semua kelas.

5. KESIMPULAN

Hasil penelitian menunjukkan bahwa pemanfaatan teknik data mining dengan menggunakan algoritma K-Means Clustering dan Random Forest mampu memberikan pendekatan yang efektif dalam mengolah dan menganalisis data kinerja guru secara sistematis, terstruktur, dan berbasis data. Pendekatan ini memungkinkan proses evaluasi kinerja yang sebelumnya cenderung subjektif menjadi lebih objektif dan terukur.

Algoritma K-Means Clustering mampu mengelompokkan data kinerja guru ke dalam beberapa cluster berdasarkan tingkat kemiripan karakteristik data, sehingga menghasilkan kategori kinerja seperti tinggi, sedang, dan rendah. Hasil pengelompokan ini memberikan gambaran yang jelas mengenai distribusi kinerja guru serta membantu pihak sekolah dalam memahami pola umum yang terdapat dalam data administrasi. Selain itu, proses ini juga mempermudah dalam mengidentifikasi kelompok guru yang memerlukan pembinaan maupun yang telah menunjukkan kinerja optimal.

Algoritma Random Forest menunjukkan kinerja yang baik dalam melakukan klasifikasi kinerja guru berdasarkan atribut-atribut yang tersedia. Melalui proses perhitungan entropy dan information gain, diketahui bahwa tidak semua atribut memiliki tingkat pengaruh yang sama dalam menentukan kelas kinerja guru. Atribut dengan nilai information gain tertinggi memiliki peran yang lebih dominan dalam proses pengambilan keputusan, sehingga dapat dijadikan sebagai indikator utama dalam evaluasi kinerja guru.

Kombinasi antara metode K-Means Clustering dan Random Forest memberikan pendekatan analisis yang

saling melengkapi. K-Means Clustering berperan dalam tahap eksplorasi data untuk menemukan pola dan struktur, sedangkan Random Forest berperan dalam proses klasifikasi yang lebih terarah dan sistematis. Integrasi kedua metode ini mampu meningkatkan kualitas analisis data serta memberikan hasil yang lebih komprehensif dalam memahami kinerja guru.

Hasil penelitian ini dapat dimanfaatkan oleh pihak sekolah sebagai dasar dalam melakukan evaluasi kinerja guru secara objektif, transparan, dan berbasis data. Informasi yang dihasilkan dapat digunakan untuk mendukung pengambilan kebijakan, seperti pembinaan guru, peningkatan kompetensi, serta perencanaan strategi pengelolaan sumber daya manusia di lingkungan sekolah. Selain itu, penelitian ini memberikan kontribusi dalam mengatasi keterbatasan pada penelitian sebelumnya yang umumnya hanya menggunakan satu metode analisis dalam mengevaluasi kinerja guru. Dengan mengombinasikan K-Means Clustering dan Random Forest, penelitian ini mampu menghadirkan pendekatan yang lebih komprehensif, tidak hanya dalam mengidentifikasi pola melalui pengelompokan data, tetapi juga dalam melakukan klasifikasi secara lebih akurat. Dengan demikian, penelitian ini turut memperkaya pengembangan metode analisis kinerja guru berbasis data mining.

6. SARAN

Penelitian ini masih memiliki beberapa keterbatasan, di antaranya jumlah data yang relatif terbatas dan hanya berasal dari satu sekolah, sehingga hasil yang diperoleh belum dapat digeneralisasikan secara luas. Selain itu, data yang digunakan masih berfokus pada data administrasi sekolah dan belum sepenuhnya mencerminkan seluruh aspek kinerja guru, seperti aspek pedagogik, profesional, sosial, dan kepribadian. Penelitian ini juga belum melakukan perbandingan secara komprehensif dengan algoritma lain yang berpotensi memberikan tingkat akurasi dan performa yang lebih optimal. Oleh karena itu, penelitian selanjutnya disarankan untuk menggunakan dataset yang lebih besar, beragam, dan berasal dari berbagai sekolah atau wilayah, serta menambahkan variabel yang lebih representatif seperti data kehadiran, hasil evaluasi pembelajaran, dan umpan balik siswa. Selain itu, perlu dilakukan perbandingan dengan berbagai algoritma lain seperti Support Vector Machine (SVM), Neural Network, Decision Tree, maupun metode ensemble lainnya, serta menggunakan teknik evaluasi yang lebih komprehensif seperti cross-validation, precision, recall, dan F1-score. Pengembangan sistem berbasis aplikasi atau dashboard interaktif juga sangat direkomendasikan agar hasil analisis dapat diimplementasikan secara langsung sebagai sistem pendukung keputusan di lingkungan sekolah. Dengan pengembangan tersebut, penelitian di masa mendatang diharapkan tidak hanya mampu meningkatkan akurasi dan kualitas model, tetapi juga memberikan dampak nyata dalam meningkatkan

efektivitas evaluasi kinerja guru, mendukung pengambilan keputusan yang lebih objektif, serta berkontribusi terhadap peningkatan kualitas pengelolaan sumber daya manusia dan mutu pendidikan secara berkelanjutan.

7. REFERENSI

- Aini, N., Arif, M., Agustin, I. T., & Toyibah, Z. B. (2024). Implementasi Algoritma Random Forest untuk Klasifikasi Bidang MSIB di Prodi Pendidikan Informatika. *Jurnal Informatika*, 11(1), 11–16. <https://doi.org/10.31294/inf.v11i1.20637>
- Anwar, N. M., Amin, S. J., & Akib, M. (2025). *Respons dan Kesiapan Guru Madrasah dalam Menghadapi Transformasi Pembelajaran Berbasis Teknologi : Studi Kualitatif di Madrasah Aliyah DDI Galla Raya Raya Program Studi Pendidikan Agama Islam , Fakultas Pascasarjana , Institut Agama Islam Negeri Pare- Tea*. 5(9), 2730–2740.
- Ariawan, I. K., Putra, I. G., & Pratama, I. P. (2023). Optimization of K-means clustering using elbow method. *Journal of Data Science*, 7(2), 101–110.
- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32.
- Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From data mining to knowledge discovery in databases. *AI Magazine*, 17(3), 37–54.
- Han, J., Kamber, M., & Pei, J. (2012). *Data mining: Concepts and techniques*. Morgan Kaufmann.
- Haris, R., Haryo, W., Wahyu Pujiharto, E., Yuza, A., Kusriani, K., & Kusnawi, K. (2024). Prediksi Banjir Di Dki Jakarta Dengan Menggunakan Algoritma K-Means Dan Random Forest. *Jurnal Informatika Dan Teknologi Komputer (J-ICOM)*, 5(1), 43–49. <https://doi.org/10.55377/j-icom.v5i1.8153>
- Jain, A. K. (2010). Data clustering: 50 years beyond K-means. *Pattern Recognition Letters*, 31(8), 651–666.
- Kusriani, & Luthfi, E. T. (2009). *Algoritma data mining*. Andi.
- Mulia Mahendra Alvanof, A. (2024). Penerapan random forest dalam deteksi ransomware. *Jurnal Teknologi Informasi*, 8(2), 120–130.
- Riani, A. P., Voutama, A., & Ridwan, T. (2023). Penerapan K-Means Clustering Dalam Pengelompokan Hasil Belajar Peserta Didik Dengan Metode Elbow. *Jurnal Teknologi Sistem Informasi Dan Sistem Komputer TGD*, 6(1), 164–172.

- Saputra, E. A., & Nataliani, Y. (2021). Analisis Pengelompokan Data Nilai Siswa untuk Menentukan Siswa Berprestasi Menggunakan Metode Clustering K-Means. *Journal of Information Systems and Informatics*, 3(3), 424–439. <https://doi.org/10.51519/journalisi.v3i3.164>
- Selatan, J. R., Yogyakarta, D. I., Selatan, J. R., Yogyakarta, D., Wafiq, M., & Ahmad, U. (2021). *PENGARUH TEKNOLOGI DALAM DUNIA PENDIDIKAN*. 18(2), 91–100. <https://doi.org/10.46781/al-mutharahah.v18i2.303>
- Sofiyatus Zawayah, Lailatul Qodriyah, & Moh. Badri Tamam. (2024). Klasifikasi Prestasi Akademik Mahasiswa Menggunakan Metode Random Forest. *Journal of Digital Business and Information Technology*, 1(2), 61–71. <https://doi.org/10.23971/jobit.v1i2.317>
- Suliman, S. (2021). Implementasi Data Mining Terhadap Prestasi Belajar Mahasiswa Berdasarkan Pergaulan dan Sosial Ekonomi Dengan Algoritma K-Means Clustering. *Simkom*, 6(1), 1–11. <https://doi.org/10.51717/simkom.v6i1.48>
- Suncaka, E. (2023). MENINJAU PERMASALAHAN RENDAHNYA KUALITAS PENDIDIKAN DI INDONESIA. *JURNAL MANAJEMEN DAN PENDIDIKAN*, 02(03), 36–49.
- Tan, P. N., Steinbach, M., & Kumar, V. (2019). Introduction to data mining. Pearson.
- Witten, I. H., Frank, E., & Hall, M. A. (2016). Data mining: Practical machine learning tools. Morgan Kaufmann.
- Yoliadi, D. N. (2023). Data Mining Dalam Analisis Tingkat Penjualan Barang Elektronik Menggunakan Algoritma K-Means. *Insearch: Information System Research Journal*, 3(01). <https://doi.org/10.15548/isrj.v3i01.5829>
- Zhang, T., Ramakrishnan, R., & Livny, M. (1996). BIRCH clustering method. *SIGMOD Record*, 25(2), 103–114.
- Zul'Aini, N. H., Lubis, I., & Ria Pasaribu, T. (2024). Implementation of Data Mining Teacher Performance Assessment Using the K-means Clustering Method in Student Learning Styles in the 4.0 Era. *Journal of Engineering, Technology and Computing (JETCom)*, 3(1), 1–21. <https://doi.org/10.63893/jetcom.v3i1.137>