

Machine Learning-Based Teacher Performance Classification Using Administrative and Credit Point Assessment (PAK) Data: A Comparative Study of Decision Tree and Naive Bayes

Nur Ifani Chaerunnisa ^{1*}, Marcelinus Yosep Teguh Sulistyono ^{2*}

*Sistem Informasi, Universitas Dian Nuswantoro, Semarang

112202206749@mhs.dinus.ac.id ¹, teguh.sulistyono@dsn.dinus.ac.id ²

Article Info

Article history:

Received 2026-01-28

Revised 2026-02-22

Accepted 2026-04-08

Keyword:

*Teacher Performance,
Machine Learning,
Classification,
Decision Tree,
Naive Bayes.*

ABSTRACT

Teacher performance evaluation plays a strategic role in improving educational quality and supporting data-driven decision making. However, conventional evaluation approaches are often subjective and lack systematic data analysis. This study proposes a machine learning-based classification model to evaluate teacher performance using administrative data and Credit Point Assessment (PAK) records. The dataset consists of 30 teacher records with attributes including rank, functional position, and PAK score. Performance categories were derived based on predefined PAK score intervals, forming a structured rule-based classification framework. Data preprocessing included label encoding and min-max normalization, followed by an 80:20 stratified train-test split. Two supervised learning algorithms, Decision Tree and Naive Bayes, were implemented and evaluated using accuracy, precision, recall, F1-score, confusion matrix, and 5-fold cross-validation to ensure robustness. Experimental results show that both models achieved an accuracy of 0.83, with Decision Tree demonstrating more stable performance and higher interpretability. The relatively high performance is influenced by the deterministic structure of the PAK-based categorization and the limited dataset size. These findings indicate that Decision Tree is effective for automating structured administrative evaluation rules, while further validation using larger and multi-institutional datasets is necessary to improve generalizability. This study contributes to the development of transparent and data-driven teacher evaluation systems in educational institutions.



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

I. PENDAHULUAN

Evaluasi kinerja guru merupakan komponen fundamental dalam upaya peningkatan mutu pendidikan. Guru memiliki peran strategis dalam menentukan kualitas proses pembelajaran dan capaian belajar peserta didik, sehingga kinerja guru menjadi faktor kunci dalam keberhasilan penyelenggaraan pendidikan [1]. Oleh karena itu, sistem evaluasi kinerja guru perlu dirancang secara objektif, sistematis, dan berbasis data agar mampu mencerminkan kondisi kinerja secara nyata dan akurat [2].

Pada praktiknya, evaluasi kinerja guru di banyak institusi pendidikan masih dilakukan menggunakan pendekatan administratif dan konvensional. Penilaian kinerja umumnya didasarkan pada pemeriksaan dokumen, laporan kegiatan, dan

pertimbangan subjektif pihak tertentu [3]. Pendekatan tersebut berpotensi menimbulkan ketidakkonsistenan hasil evaluasi serta kesulitan dalam mengidentifikasi pola kinerja guru secara menyeluruh [4]. Akibatnya, hasil evaluasi sering kali kurang optimal dalam mendukung pengambilan keputusan strategis dan pengembangan profesional guru.

Perkembangan teknologi informasi membuka peluang untuk mengatasi keterbatasan tersebut melalui pendekatan berbasis data. Salah satu pendekatan yang banyak digunakan adalah machine learning, yang memungkinkan sistem mempelajari pola dari data historis dan menghasilkan prediksi atau klasifikasi secara otomatis [5]. Dalam bidang pendidikan, machine learning telah dimanfaatkan untuk menganalisis performa peserta didik, memprediksi hasil

belajar, serta mendukung sistem evaluasi yang lebih objektif dan terukur [6], [7].

Data administratif guru, seperti data kepegawaian dan Penetapan Angka Kredit (PAK), merupakan sumber data terstruktur yang memiliki potensi besar untuk dianalisis menggunakan teknik machine learning [8], [9]. Data PAK merepresentasikan aktivitas profesional guru secara kuantitatif, mencakup kegiatan pembelajaran, pengembangan keprofesian berkelanjutan, dan tugas penunjang lainnya. Apabila data tersebut dianalisis secara sistematis, maka pola kinerja guru dapat diidentifikasi secara lebih objektif dan berbasis bukti [10].

Sejumlah penelitian terdahulu menunjukkan bahwa teknik klasifikasi dalam machine learning efektif digunakan untuk analisis data pendidikan. Metode klasifikasi mampu mengelompokkan objek ke dalam kategori tertentu berdasarkan karakteristik yang dimiliki, sehingga sesuai untuk digunakan dalam evaluasi kinerja [11], [12]. Namun demikian, performa model klasifikasi sangat bergantung pada karakteristik data serta algoritma yang digunakan [13].

Di antara berbagai algoritma klasifikasi, Decision Tree dan Naive Bayes merupakan dua metode yang sering digunakan karena kesederhanaan dan kemudahan interpretasinya [14]. Decision Tree memiliki keunggulan dalam membentuk aturan keputusan yang jelas dan mudah dipahami, sehingga sangat sesuai untuk konteks pendidikan yang membutuhkan transparansi hasil analisis [15], [16]. Sementara itu, Naive Bayes menggunakan pendekatan probabilistik berdasarkan Teorema Bayes dengan asumsi independensi antar fitur, serta sering digunakan sebagai model pembanding karena efisiensi komputasinya [4], [17].

Meskipun kedua algoritma tersebut banyak digunakan, penelitian yang secara khusus membandingkan kinerja Decision Tree dan Naive Bayes dalam konteks klasifikasi kinerja guru berbasis data administratif masih relatif terbatas [18]. Sebagian besar penelitian sebelumnya lebih berfokus pada analisis kinerja peserta didik atau prediksi hasil belajar, sehingga masih terdapat celah penelitian (research gap) dalam penerapan machine learning untuk evaluasi kinerja guru berbasis data kepegawaian [19].

Berdasarkan latar belakang tersebut, penelitian ini mengusulkan pendekatan klasifikasi berbasis machine learning untuk mengevaluasi kinerja guru menggunakan data pegawai dan Penetapan Angka Kredit (PAK) [20]. Dua algoritma klasifikasi, yaitu Decision Tree dan Naive Bayes, diterapkan dan dibandingkan untuk menilai performanya dalam mengklasifikasikan kinerja guru [21].

Tujuan penelitian ini adalah: (1) menerapkan algoritma Decision Tree dan Naive Bayes dalam klasifikasi kinerja guru, (2) membandingkan performa kedua algoritma menggunakan metrik evaluasi klasifikasi, dan (3) menentukan model klasifikasi yang paling sesuai untuk mendukung evaluasi kinerja guru berbasis data.

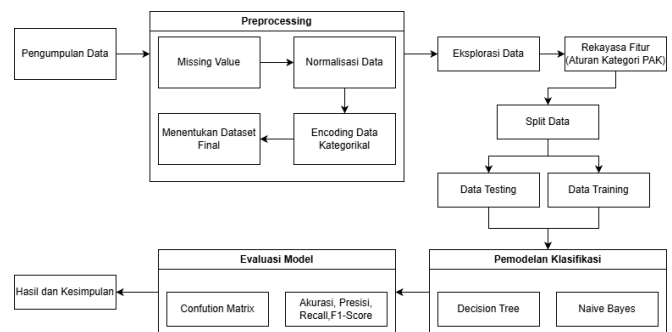
Hasil penelitian ini diharapkan dapat memberikan kontribusi dalam pengembangan sistem evaluasi kinerja guru yang lebih objektif, transparan, dan berbasis data. Selain itu, penelitian ini diharapkan dapat menjadi referensi bagi institusi pendidikan dalam memanfaatkan machine learning sebagai alat bantu pengambilan keputusan dalam manajemen kinerja pendidik.

II. METODE

A. Desain Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan metode machine learning, khususnya teknik klasifikasi. Pendekatan kuantitatif dipilih karena penelitian berfokus pada pengolahan data numerik dan kategorikal yang dianalisis secara matematis untuk menemukan pola dan hubungan antar atribut. Teknik klasifikasi digunakan untuk mengelompokkan kinerja guru ke dalam kategori tertentu berdasarkan karakteristik administratif dan nilai Penetapan Angka Kredit (PAK).

Pendekatan machine learning diharapkan mampu memberikan hasil klasifikasi kinerja guru yang lebih objektif, terukur, dan konsisten dibandingkan metode evaluasi konvensional yang cenderung bersifat subjektif.



Gambar 1. Metodologi.

B. Pengumpulan dan Integrasi Data

Pengumpulan data dilakukan menggunakan metode dokumentasi, yaitu dengan menghimpun data administratif guru yang bersumber dari dokumen resmi SMP Negeri 1 Gemuh. Data yang digunakan merupakan data sekunder dan data riil, yang mencerminkan kondisi aktual kinerja guru pada periode pengamatan tertentu, bukan data simulasi maupun data buatan.

Data yang dikumpulkan terdiri atas data pegawai guru dan data Penetapan Angka Kredit (PAK). Data pegawai memuat informasi golongan, jabatan fungsional, dan unit kerja, sedangkan data PAK merepresentasikan capaian kinerja guru secara kuantitatif. Seluruh data kemudian diintegrasikan berdasarkan identitas guru sehingga membentuk satu dataset yang utuh dan konsisten.

Untuk menjaga etika penelitian, identitas pribadi seperti Nomor Induk Pegawai (NIP) disamarkan. Dataset akhir yang digunakan dalam penelitian ini terdiri dari 30 data guru

dengan delapan atribut utama. Ringkasan dataset ditampilkan pada Tabel 1, sedangkan seluruh proses analisis dan pemodelan dilakukan menggunakan dataset lengkap.

TABEL I
DATASET KINERJA GURU SMP NEGERI 1 GEMUH (RINGKASAN DATA ASLI)

No	NIP	Kode Pegawai	Golongan	Jabatan Fungsional	Nilai PAK	Kategori Kinerja
1	1968xxxx1993	PG-01	IV/b	Guru Ahli Madya	525	Sangat Baik
2	1968xxxx1994	PG-02	IV/a	Guru Ahli Madya	480	Sangat Baik
3	1968xxxx1995	PG-03	IV/a	Guru Ahli Madya	455	Baik
*	*****	*****	*****	*****	***	***
*	*****	*****	*****	*****	***	***
28	1968xxxx2020	PG-28	III/a	Guru Pertama	250	Kurang
29	1968xxxx2021	PG-29	II/d	Guru Pertama	240	Kurang

Catatan: Tabel ditampilkan dalam bentuk ringkasan. Seluruh proses preprocessing, pemodelan, dan evaluasi dilakukan menggunakan dataset lengkap.

1) *Distribusi Dataset*

TABEL II
DISTRIBUSI KATEGORI KINERJA GURU

Kategori Kinerja	Jumlah Data
Sangat Baik	4
Baik	6
Cukup	10
Kurang	10
Total	30

Berdasarkan distribusi dataset, kategori Cukup dan Kurang memiliki jumlah data yang relatif lebih banyak dibandingkan kategori Sangat Baik. Ketidakseimbangan jumlah data pada beberapa kategori berpotensi mempengaruhi sensitivitas model terhadap kelas minoritas dan perlu diperhatikan dalam interpretasi hasil klasifikasi.

C. *Preprocessing Data*

Preprocessing data dilakukan untuk memastikan bahwa dataset berada dalam kondisi siap digunakan pada proses pemodelan klasifikasi. Tahapan preprocessing meliputi penanganan missing value, normalisasi data, dan encoding data kategorikal.

1) *Penanganan Missing Value* dilakukan terhadap seluruh atribut dataset. Hasil pemeriksaan menunjukkan bahwa tidak terdapat missing value, sehingga seluruh data dapat digunakan secara langsung. Secara metodologis, apabila ditemukan missing value pada atribut numerik, maka dapat digunakan metode imputasi rata-rata dengan rumus:

$$\bar{x} = \frac{\sum_{i=1}^n xi}{n}$$

dengan:

\bar{x} = nilai rata-rata atribut

xi = nilai data ke- i

n = jumlah data tanpa missing value

Kategori kinerja guru pada dataset ini terdiri atas Sangat Baik, Baik, Cukup, dan Kurang. Kategori tersebut berperan sebagai label kelas (class label) dalam proses klasifikasi menggunakan algoritma Decision Tree dan Naive Bayes. Sementara itu, atribut golongan dan jabatan fungsional mencerminkan jenjang karier guru, dan atribut nilai PAK berfungsi sebagai indikator kuantitatif utama yang merepresentasikan aktivitas dan pencapaian kinerja guru.

2) *Normalisasi Data* dilakukan untuk menyeragamkan skala nilai numerik agar tidak terjadi dominasi atribut tertentu dalam proses pembelajaran model. Pada penelitian ini, normalisasi difokuskan pada atribut nilai PAK, karena memiliki rentang nilai yang lebih besar dibandingkan atribut lainnya.

Metode normalisasi yang digunakan adalah Min–Max Normalization, yang mentransformasikan nilai data ke dalam rentang 0 hingga 1 dengan rumus:

$$x' = \frac{x - x_{min}}{x_{max} - x_{min}}$$

Keterangan:

x' = nilai hasil normalisasi

x = nilai asli

x_{min} = nilai minimum atribut

x_{max} = nilai maksimum atribut

3) *Encoding Data Kategorikal*. Atribut kategorikal seperti golongan, jabatan fungsional, unit kerja, dan kategori kinerja guru tidak dapat diproses secara langsung oleh algoritma machine learning. Oleh karena itu, dilakukan proses Label Encoding untuk mengubah data kategorikal ke dalam bentuk numerik.

Label Encoding dipilih karena sesuai dengan karakteristik atribut serta algoritma yang digunakan dalam penelitian ini, khususnya Decision Tree dan Naive Bayes.

$$f = K \rightarrow \mathbb{Z}$$

Keterangan:

K = himpunan kategori

\mathbb{Z} = bilangan bulat hasil pemetaan

4) *Penentuan Dataset Final* Setelah seluruh tahapan preprocessing dilakukan, diperoleh dataset final yang digunakan dalam penelitian ini. Dataset final terdiri dari 30 data guru yang telah melalui proses pembersihan data,

normalisasi, dan encoding, sehingga seluruh atribut berada dalam bentuk numerik dan siap digunakan dalam proses pemodelan klasifikasi.

Atribut yang digunakan sebagai fitur (input) pada dataset final meliputi golongan, jabatan fungsional, unit kerja, dan nilai PAK. Sementara itu, kategori kinerja guru ditetapkan sebagai variabel target (label kelas) yang dibentuk melalui proses rekayasa fitur berdasarkan nilai PAK.

Penentuan label kategori kinerja guru dilakukan menggunakan aturan kategorisasi nilai PAK yang dirumuskan secara matematis pada Subbab Rekayasa Fitur (Aturan Kategori PAK), sehingga setiap data guru memiliki label kinerja yang jelas dan tidak ambigu.

Atribut identitas seperti Nomor Induk Pegawai (NIP) dan kode pegawai tidak disertakan dalam dataset final karena tidak memiliki pengaruh terhadap proses klasifikasi dan hanya berfungsi sebagai penanda administratif.

D. Eksplorasi Data

Eksplorasi data dilakukan untuk memperoleh gambaran awal mengenai karakteristik dataset setelah melalui tahap preprocessing. Tahap ini bertujuan untuk memastikan bahwa data memiliki variasi nilai dan distribusi kelas yang memadai sebelum dilakukan pemodelan.

Hasil eksplorasi menunjukkan bahwa nilai PAK berada pada rentang 230–540 dan mencakup seluruh kategori kinerja guru, yaitu *Sangat Baik*, *Baik*, *Cukup*, dan *Kurang*. Kondisi ini menunjukkan bahwa dataset representatif dan layak digunakan dalam proses klasifikasi.

E. Rekayasa Fitur (Aturan Kategori PAK)

Kategori kinerja guru pada penelitian ini ditentukan berdasarkan interval nilai PAK yang telah ditetapkan. Dengan demikian, proses pembentukan label kelas bersifat rule-based berdasarkan ambang batas nilai PAK. Oleh karena itu, model klasifikasi pada penelitian ini berfungsi untuk mengevaluasi kemampuan algoritma dalam merepresentasikan aturan administratif tersebut secara otomatis, bukan untuk menemukan pola laten yang tidak terstruktur.

Rekayasa fitur dilakukan dengan mengonversi nilai PAK yang bersifat numerik ke dalam kategori kinerja guru. Proses ini bertujuan untuk membentuk variabel target yang sesuai dengan pendekatan klasifikasi dan memudahkan interpretasi hasil model.

Aturan kategorisasi ditetapkan berdasarkan distribusi nilai PAK pada dataset serta praktik penilaian kinerja guru, sebagai berikut:

$$f(x) = \begin{cases} \text{Sangat Baik,} & x \geq 480 \\ \text{Baik,} & 380 \leq x < 480 \\ \text{Cukup,} & 300 \leq \text{PAK} < 380 \\ \text{Kurang,} & x < 300 \end{cases}$$

x = nilai Penetapan Angka Kredit (PAK) guru
 $f(x)$ = kategori kinerja guru (label kelas)

F. Pembagian Data

Dataset final yang telah melalui seluruh tahapan preprocessing selanjutnya dibagi menjadi data latih (training set) dan data uji (testing set). Pembagian data ini bertujuan untuk melatih model klasifikasi serta mengevaluasi kemampuan model dalam melakukan generalisasi terhadap data yang belum pernah dilihat sebelumnya.

Pada penelitian ini, proporsi pembagian data yang digunakan adalah 80% data latih dan 20% data uji. Proporsi tersebut dipilih karena umum digunakan dalam penelitian machine learning dan dinilai cukup representatif untuk melatih model sekaligus menguji performanya, terutama pada dataset dengan jumlah data terbatas.

Secara matematis, pembagian dataset dinyatakan sebagai berikut:

$$D_{total} = D_{train} \cup D_{test}$$

dengan proporsi:

$$D_{train} = 0,8 \times D_{total}, D_{test} = 0,2 \times D_{total}$$

Keterangan:

D_{total} = keseluruhan dataset final

D_{train} = data latih

D_{test} = data uji

D = jumlah data dalam himpunan D

Pembagian data dilakukan secara acak (random split) untuk menghindari bias urutan data dan memastikan bahwa data latih dan data uji memiliki karakteristik yang sebanding. Selain itu, pembagian data dilakukan menggunakan pendekatan stratified random sampling untuk menjaga proporsi distribusi kelas pada data latih dan data uji tetap seimbang sesuai dengan distribusi dataset awal.

G. Pemodelan Klasifikasi

Pemodelan klasifikasi dilakukan menggunakan dua algoritma supervised learning, yaitu Decision Tree dan Naive Bayes. Kedua algoritma digunakan untuk membandingkan pendekatan klasifikasi berbasis aturan dan probabilistik.

1) *Decision Tree* membangun model klasifikasi dalam bentuk struktur pohon keputusan dengan cara melakukan pemisahan data secara bertahap berdasarkan atribut yang paling informatif. Pemilihan atribut pada setiap node dilakukan menggunakan kriteria entropy dan information gain.

$$Entropy(S) = - \sum_{i=1}^n p_i \log_2(p_i)$$

$$Gain(S, A) = Entropy(S) - \sum_{v \in Values(A)} \left| \frac{S_v}{S} \right| Entropy(S_v)$$

Keterangan:

S = himpunan data pada node

c = jumlah kelas

p_i = proporsi data kelas ke- i

A = atribut pemisah

S_v = subset data berdasarkan nilai atribut A

Atribut dengan nilai information gain tertinggi dipilih sebagai node pemisah.

Untuk menjaga kompleksitas model dan meminimalkan risiko overfitting pada dataset yang relatif terbatas, kedalaman maksimum pohon keputusan (max_depth) dibatasi hingga tiga tingkat.

2) *Naive Bayes* merupakan algoritma klasifikasi berbasis probabilitas yang menggunakan Teorema Bayes dengan asumsi independensi antar atribut. Prinsip dasar Naive Bayes dirumuskan sebagai:

$$P(C|X) = \frac{P(X|C) \cdot P(C)}{P(X)}$$

Keterangan:

C = kelas kinerja guru

X = data input

$P(C|X)$ = probabilitas data termasuk ke kelas C

H. Evaluasi Model

Evaluasi model dilakukan untuk mengukur kemampuan algoritma klasifikasi dalam memprediksi kategori kinerja guru secara akurat dan konsisten. Proses evaluasi bertujuan untuk menilai sejauh mana model yang dibangun mampu mengklasifikasikan data uji dengan benar berdasarkan pola yang telah dipelajari dari data latih.

Pada penelitian ini, evaluasi model dilakukan menggunakan confusion matrix sebagai dasar perhitungan metrik evaluasi. Confusion matrix memberikan gambaran mengenai jumlah prediksi yang benar dan salah pada setiap kategori kinerja guru, sehingga memungkinkan analisis performa model secara lebih terstruktur.

Berdasarkan confusion matrix, digunakan beberapa metrik evaluasi yang umum dalam penelitian klasifikasi, yaitu akurasi, presisi, recall, dan F1-score. Metrik akurasi digunakan untuk mengukur tingkat ketepatan model secara keseluruhan dalam mengklasifikasikan data uji. Namun, akurasi saja tidak selalu cukup untuk menggambarkan

performa model, terutama ketika distribusi kelas tidak sepenuhnya seimbang.

Oleh karena itu, metrik presisi dan recall turut digunakan untuk mengevaluasi performa model pada masing-masing kategori kinerja guru. Presisi mengukur tingkat ketepatan prediksi model terhadap suatu kelas tertentu, sedangkan recall mengukur kemampuan model dalam mengenali seluruh data yang benar-benar termasuk ke dalam kelas tersebut.

Sebagai metrik gabungan, F1-score digunakan untuk menyeimbangkan nilai presisi dan recall, sehingga memberikan gambaran performa model yang lebih komprehensif. Penggunaan keempat metrik ini memungkinkan evaluasi model dilakukan secara menyeluruh dan objektif, tanpa bergantung pada satu ukuran kinerja saja.

Secara matematis, metrik evaluasi yang digunakan dirumuskan sebagai berikut:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

Keterangan:

TP = True Positive

TN = True Negative

FP = False Positive

FN = False Negative

Nilai-nilai metrik evaluasi tersebut selanjutnya digunakan sebagai dasar dalam membandingkan performa algoritma Decision Tree dan Naive Bayes pada tahap hasil dan pembahasan.

I. Validasi Model

Untuk mengurangi risiko overfitting dan meningkatkan reliabilitas model, dilakukan validasi silang menggunakan metode K-Fold Cross Validation dengan nilai $k = 5$. Metode ini membagi dataset menjadi lima subset, di mana setiap subset secara bergantian digunakan sebagai data uji dan sisanya sebagai data latih. Nilai performa akhir diperoleh dari rata-rata hasil pengujian pada seluruh fold. Hasil 5-fold cross-validation menunjukkan bahwa performa model relatif konsisten pada berbagai fold pengujian, yang mengindikasikan stabilitas model terhadap variasi pembagian data. Dengan demikian, validasi silang memastikan bahwa model tidak hanya menyesuaikan diri terhadap satu skenario pengujian (hold-out testing), sehingga hasil evaluasi menjadi

lebih robust dan mampu meminimalkan risiko overfitting, terutama pada dataset dengan jumlah data yang terbatas.

III. HASIL DAN PEMBAHASAN

A. Analisis Eksploratif Data

1) *Statistik Deskriptif Data* digunakan untuk memberikan gambaran umum mengenai karakteristik dataset final yang digunakan dalam penelitian ini. Analisis ini mencakup atribut numerik utama, yaitu Nilai Penetapan Angka Kredit (PAK), Golongan (ordinal), dan Jabatan Fungsional (ordinal), sebagaimana ditunjukkan pada hasil statistik deskriptif.

TABEL III
STATISTIK DESKRIPTIF DATA

Index	Nilai PAK	Golongan Ordinal	Jabatan Ordinal
count	30.0	30.0	30.0
mean	338.5	10.733333333333333	1.2666666666666666
std	94.45077170601661	1.7991057804352142	0.44977644510880355
min	225.0	8.0	1.0
25%	261.25	9.25	1.0
50%	312.5	11.0	1.0
75%	392.5	12.0	1.75
max	540.0	14.0	2.0

Berdasarkan hasil analisis, jumlah data yang digunakan dalam penelitian ini adalah 30 data guru. Nilai PAK memiliki nilai minimum sebesar 225 dan nilai maksimum sebesar 540, dengan nilai rata-rata (mean) sebesar 338,5 dan simpangan baku (standard deviation) sebesar 94,45. Nilai simpangan baku yang relatif besar menunjukkan adanya variasi nilai PAK yang cukup lebar di antara guru, yang mencerminkan perbedaan tingkat capaian kinerja.

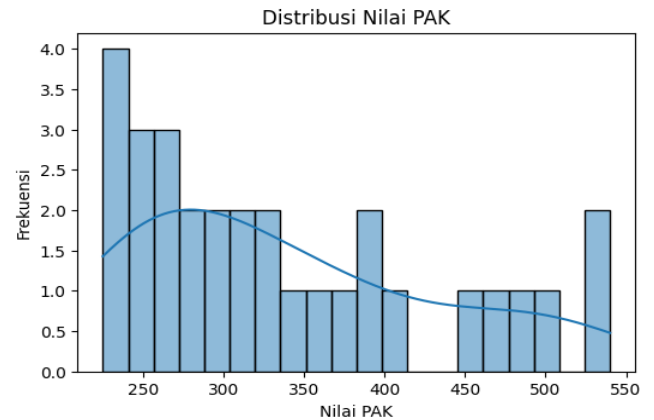
Atribut Golongan (ordinal) memiliki nilai rata-rata sebesar 10,73, dengan nilai minimum 8 dan maksimum 14. Hal ini menunjukkan bahwa data guru yang dianalisis mencakup rentang golongan yang cukup beragam, sehingga dapat memberikan kontribusi informasi yang relevan dalam proses klasifikasi kinerja guru.

Sementara itu, atribut Jabatan Fungsional (ordinal) memiliki nilai rata-rata sebesar 1,27, dengan nilai minimum 1 dan maksimum 2. Nilai simpangan baku yang relatif kecil (0,45) menunjukkan bahwa variasi jabatan fungsional pada dataset ini lebih terbatas dibandingkan dengan atribut nilai PAK dan golongan.

Secara keseluruhan, hasil statistik deskriptif menunjukkan bahwa dataset memiliki variasi nilai yang memadai, terutama pada atribut nilai PAK dan golongan, sehingga dataset dinilai representatif dan layak digunakan dalam proses klasifikasi kinerja guru.

Distribusi Nilai Penetapan Angka Kredit (PAK) dianalisis untuk melihat sebaran data serta kecenderungan nilai kinerja guru pada dataset yang digunakan. Gambar distribusi nilai PAK menunjukkan bahwa data tersebar pada rentang 225

hingga 540, dengan kepadatan data yang lebih tinggi pada nilai PAK menengah.



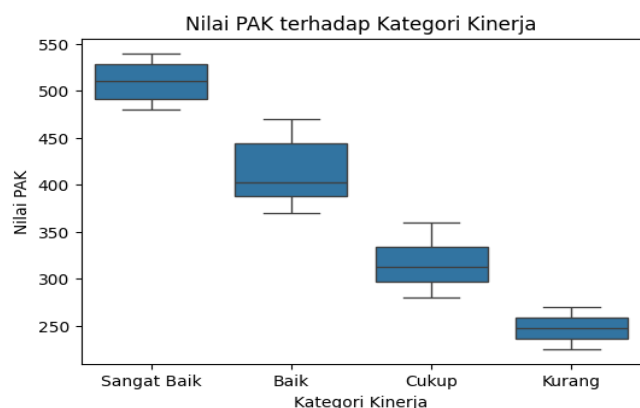
Gambar 2. Hasil Distribusi Nilai Penetapan Angka Kredit (PAK).

Berdasarkan distribusi tersebut, sebagian besar nilai PAK terkonsentrasi pada kisaran 250 hingga 350, yang mengindikasikan bahwa mayoritas guru berada pada tingkat kinerja menengah. Pada rentang ini, frekuensi data relatif lebih tinggi dibandingkan dengan rentang nilai lainnya. Sementara itu, jumlah guru dengan nilai PAK yang sangat tinggi (di atas 450) relatif lebih sedikit, yang tercermin dari frekuensi data yang semakin menurun pada bagian ekor kanan distribusi.

Distribusi nilai PAK cenderung tidak simetris sempurna dan menunjukkan kecenderungan *right-skewed* (condong ke kanan), yang menandakan adanya sejumlah kecil guru dengan nilai PAK yang sangat tinggi dibandingkan dengan mayoritas data. Pola distribusi seperti ini umum ditemukan pada data kinerja, di mana hanya sebagian kecil individu yang mencapai capaian sangat tinggi.

Sebaran nilai PAK yang cukup lebar dan tidak terpusat pada satu rentang tertentu menunjukkan bahwa dataset memiliki variasi yang memadai. Kondisi ini mendukung proses klasifikasi karena model memiliki cukup informasi untuk membedakan kategori kinerja guru berdasarkan perbedaan nilai PAK.

2) *Hubungan Nilai PAK terhadap Kategori Kinerja Guru* Analisis hubungan antara nilai Penetapan Angka Kredit (PAK) dan kategori kinerja guru dilakukan untuk melihat bagaimana perbedaan nilai PAK terdistribusi pada setiap kategori kinerja. Hubungan ini divisualisasikan menggunakan diagram boxplot, yang menggambarkan sebaran nilai, median, serta variasi data pada masing-masing kategori kinerja.



Gambar 3. Hasil Hubungan Nilai PAK terhadap Kategori Kinerja Guru.

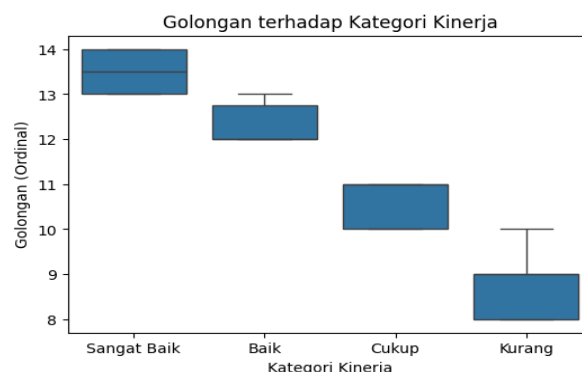
Berdasarkan hasil analisis, terlihat bahwa setiap kategori kinerja memiliki rentang nilai PAK yang berbeda secara jelas. Kategori Sangat Baik menunjukkan nilai PAK yang paling tinggi, dengan median berada di atas kategori lainnya serta rentang nilai yang relatif sempit. Hal ini mengindikasikan bahwa guru dengan kategori kinerja sangat baik memiliki capaian PAK yang konsisten tinggi.

Pada kategori Baik, nilai PAK berada pada rentang menengah atas dengan variasi yang lebih lebar dibandingkan kategori sangat baik. Meskipun terdapat tumpang tindih nilai dengan kategori di atasnya, median nilai PAK pada kategori ini tetap lebih rendah, yang menunjukkan adanya perbedaan tingkat kinerja yang masih dapat dibedakan secara kuantitatif.

Kategori Cukup memiliki nilai PAK yang lebih rendah dengan sebaran data yang relatif sempit. Median nilai PAK pada kategori ini berada di bawah kategori baik, yang mencerminkan penurunan capaian kinerja guru secara bertahap. Sementara itu, kategori Kurang menunjukkan nilai PAK terendah dengan rentang nilai yang paling kecil, menandakan bahwa guru dalam kategori ini memiliki capaian PAK yang relatif homogen pada tingkat rendah.

Secara keseluruhan, hasil ini menunjukkan bahwa nilai PAK memiliki hubungan yang kuat dan konsisten dengan kategori kinerja guru. Pola sebaran yang berjenjang dari kategori Sangat Baik hingga Kurang menguatkan validitas aturan kategorisasi nilai PAK yang digunakan dalam penelitian ini serta menunjukkan bahwa nilai PAK merupakan atribut numerik utama dalam membedakan kinerja guru.

3) *Hubungan Golongan terhadap Kategori Kinerja Guru* Analisis hubungan antara golongan dan kategori kinerja guru dilakukan untuk melihat bagaimana tingkat golongan berkontribusi terhadap pengelompokan kinerja guru. Golongan pada penelitian ini diperlakukan sebagai data ordinal, sehingga perbedaannya merepresentasikan jenjang kepangkatan guru secara berurutan.



Gambar 4. Hasil Hubungan Golongan terhadap Kategori Kinerja Guru

Berdasarkan hasil visualisasi, terlihat bahwa guru dengan kategori kinerja Sangat Baik cenderung berada pada golongan yang lebih tinggi dibandingkan kategori lainnya. Median golongan pada kategori ini berada pada tingkat tertinggi, dengan rentang nilai yang relatif sempit. Hal ini menunjukkan bahwa guru dengan kinerja sangat baik umumnya telah mencapai jenjang kepangkatan yang lebih tinggi dan relatif homogen.

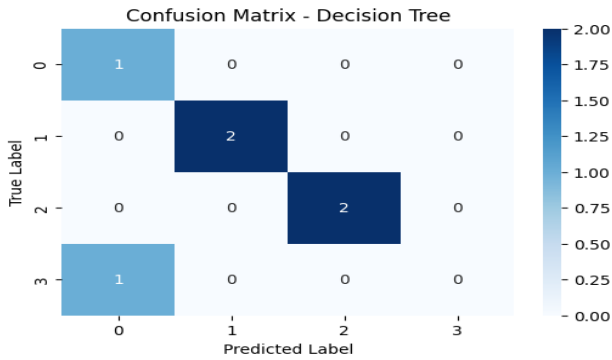
Pada kategori Baik, golongan guru berada pada tingkat menengah atas dengan variasi yang sedikit lebih besar. Meskipun sebagian guru pada kategori ini memiliki golongan yang mendekati kategori sangat baik, median golongannya tetap lebih rendah. Kondisi ini menunjukkan adanya perbedaan jenjang kepangkatan yang masih dapat dibedakan antar kategori kinerja.

Kategori Cukup menunjukkan nilai golongan yang lebih rendah dan relatif terpusat pada rentang tertentu. Hal ini mengindikasikan bahwa guru dengan kinerja cukup umumnya berada pada golongan menengah. Sementara itu, kategori Kurang didominasi oleh golongan yang lebih rendah dengan sebaran yang terbatas, mencerminkan bahwa guru dengan kinerja kurang cenderung berada pada jenjang kepangkatan awal.

Secara keseluruhan, hasil ini menunjukkan adanya kecenderungan hubungan positif antara golongan dan kategori kinerja guru, di mana peningkatan golongan diikuti oleh peningkatan kategori kinerja. Namun demikian, variasi nilai pada beberapa kategori mengindikasikan bahwa golongan bukan satu-satunya faktor penentu kinerja guru, sehingga diperlukan kombinasi atribut lain dalam proses klasifikasi.

B. Hasil Pemodelan Klasifikasi

1) *Hasil Klasifikasi Menggunakan Decision Tree* dievaluasi menggunakan data uji dengan bantuan confusion matrix. Confusion matrix digunakan untuk melihat kesesuaian antara label aktual dan label hasil prediksi yang dihasilkan oleh model.



Gambar 5 Hasil Klasifikasi Menggunakan Decision Tree

Berdasarkan confusion matrix yang diperoleh, terlihat bahwa sebagian besar data uji berhasil diklasifikasikan dengan benar oleh model Decision Tree. Hal ini ditunjukkan oleh nilai-nilai yang dominan berada pada diagonal utama matriks, yang merepresentasikan prediksi yang sesuai antara kelas aktual dan kelas prediksi.

Sebagian besar data uji berhasil diklasifikasikan dengan benar, meskipun masih terdapat kesalahan pada kategori dengan jumlah data yang sangat terbatas, yang menunjukkan bahwa Decision Tree mampu mempelajari pola hubungan antara atribut input dan kategori kinerja secara cukup baik pada sebagian besar kategori. Hasil ini mengindikasikan bahwa model Decision Tree memiliki kemampuan yang cukup baik dalam membedakan sebagian besar kategori kinerja guru berdasarkan data yang digunakan.

Secara keseluruhan, hasil evaluasi menunjukkan bahwa model Decision Tree menunjukkan performa klasifikasi yang baik dan relatif stabil pada dataset penelitian ini, sehingga layak dipertimbangkan sebagai model utama dalam klasifikasi kinerja guru.

2) *Kinerja Model Decision Tree Berdasarkan Classification Report* Evaluasi performa model Decision Tree dilakukan menggunakan classification report untuk melihat kinerja model pada setiap kategori kinerja guru secara lebih rinci. Classification report menyajikan nilai presisi, recall, dan F1-score untuk masing-masing kelas, sehingga memberikan gambaran yang lebih komprehensif dibandingkan dengan akurasi secara keseluruhan.

TABEL IV
CLASSIFICATION REPORT MODEL DECISION TREE

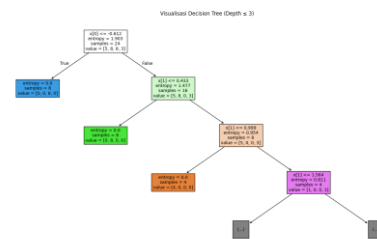
Kategori Kinerja	Precision	Recall	F1-Score	Support
Kelas 0	0.5	1	0.67	1
Kelas 1	1	1	1	2
Kelas 2	1	1	1	2
Kelas 3	0	0	0	1
Accuracy	—	—	0.83	6
Macro Average	0.62	0.75	0.67	6
Weighted Average	0.75	0.83	0.78	6

Berdasarkan hasil classification report, model Decision Tree menunjukkan performa yang sangat baik pada kelas 1 dan kelas 2, dengan nilai presisi, recall, dan F1-score masing-masing sebesar 1,00. Hal ini menunjukkan bahwa model mampu mengenali dan mengklasifikasikan kategori kinerja tersebut secara konsisten tanpa kesalahan pada data uji.

Pada kelas 0, model memiliki nilai recall sebesar 1,00, yang menunjukkan bahwa seluruh data aktual pada kelas tersebut berhasil dikenali. Namun, nilai presisi sebesar 0,50 mengindikasikan bahwa terdapat prediksi dari kelas lain yang keliru diklasifikasikan sebagai kelas 0. Kondisi ini menunjukkan bahwa meskipun model mampu menangkap seluruh data aktual kelas 0, tingkat ketepatan prediksi untuk kelas tersebut masih terbatas.

Sementara itu, pada kelas 3, model Decision Tree tidak berhasil melakukan klasifikasi dengan baik, yang ditunjukkan oleh nilai presisi, recall, dan F1-score sebesar 0,00. Hal ini disebabkan oleh jumlah data yang sangat terbatas pada kelas tersebut, sehingga model kesulitan mempelajari pola yang representatif.

Secara keseluruhan, model Decision Tree menghasilkan nilai akurasi sebesar 0,83, dengan nilai macro average F1-score sebesar 0,67 dan weighted average F1-score sebesar 0,78. Hasil ini menunjukkan bahwa model memiliki performa yang cukup baik secara umum, namun masih memiliki keterbatasan dalam mengklasifikasikan kategori kinerja dengan jumlah data yang sangat sedikit.

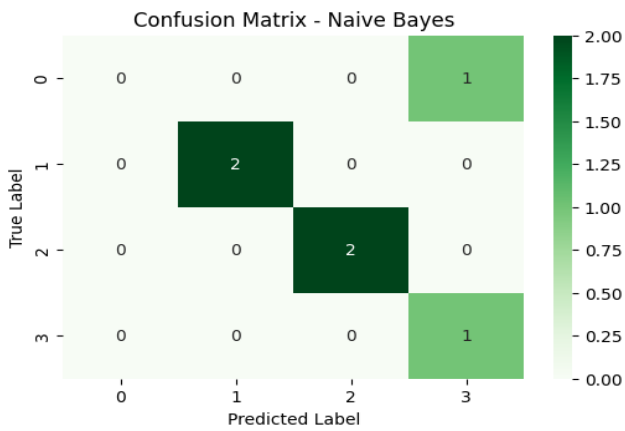


Gambar 6. Visualisasi Struktur Pohon Keputusan Decision Tree (Depth ≤ 3)

Gambar di atas menunjukkan struktur pohon keputusan yang dihasilkan oleh algoritma Decision Tree dengan kedalaman maksimum tiga tingkat. Atribut yang menjadi pemisah utama pada root node menunjukkan kontribusi informasi terbesar dalam proses klasifikasi, yang dalam penelitian ini berkaitan dengan nilai PAK. Struktur percabangan yang relatif sederhana menunjukkan bahwa model tidak membentuk aturan yang terlalu kompleks. Node dengan entropy bernilai 0 menunjukkan bahwa data pada cabang tersebut telah terklasifikasi secara konsisten dalam satu kategori.

3) *Hasil Klasifikasi Menggunakan Naive Bayes* Confusion matrix digunakan untuk mengevaluasi kinerja model Naive Bayes dengan membandingkan label aktual dan

label hasil prediksi pada data uji. Melalui confusion matrix, dapat diamati secara langsung pola kesalahan dan keberhasilan klasifikasi pada masing-masing kategori kinerja guru.



Gambar 7 Hasil Confusion Matrix Naive Bayes

Berdasarkan confusion matrix yang diperoleh, model Naive Bayes menunjukkan performa yang berbeda pada setiap kategori kinerja. Pada kelas 1 dan kelas 2, seluruh data uji berhasil diklasifikasikan dengan benar. Hal ini ditunjukkan oleh nilai prediksi yang sepenuhnya berada pada diagonal utama matriks, yang mengindikasikan tidak adanya kesalahan klasifikasi pada kedua kelas tersebut.

Sebaliknya, pada kelas 0, model Naive Bayes tidak berhasil melakukan prediksi yang benar. Seluruh data aktual pada kelas ini diprediksi sebagai kelas lain, yang menunjukkan bahwa model mengalami kesulitan dalam mengenali pola pada kategori kinerja dengan jumlah data yang sangat terbatas.

Pada kelas 3, model berhasil mengenali data aktual dengan baik, yang ditunjukkan oleh nilai prediksi benar pada diagonal. Namun demikian, masih terdapat prediksi keliru dari kelas lain yang masuk ke dalam kelas ini, sehingga menyebabkan tingkat ketepatan prediksi (presisi) pada kelas 3 menjadi lebih rendah.

Secara keseluruhan, confusion matrix menunjukkan bahwa model Naive Bayes memiliki kemampuan klasifikasi yang baik pada kategori kinerja dengan jumlah data yang relatif lebih banyak, namun performanya menurun pada kategori dengan jumlah data yang sangat sedikit. Hasil ini mengindikasikan bahwa distribusi data dan asumsi independensi antar atribut berpengaruh terhadap kinerja model Naive Bayes dalam penelitian ini.

4) *Kinerja Model Naive Bayes Berdasarkan Classification Report* Classification report digunakan untuk mengevaluasi performa model Naive Bayes secara lebih rinci pada setiap kategori kinerja guru. Melalui classification report, dapat diketahui nilai presisi, recall, F1-score, serta jumlah data (support) pada masing-masing kelas, sehingga memberikan gambaran yang lebih komprehensif dibandingkan akurasi secara keseluruhan.

TABEL V
CLASSIFICATION REPORT MODEL NAIVE BAYES

Kategori Kinerja	Precision	Recall	F1-Score	Support
Kelas 0	0	0	0	1
Kelas 1	1	1	1	2
Kelas 2	1	1	1	2
Kelas 3	0.5	1	0.67	1
Accuracy	—	—	0.83	6
Macro Average	0.62	0.75	0.67	6
Weighted Average	0.75	0.83	0.78	6

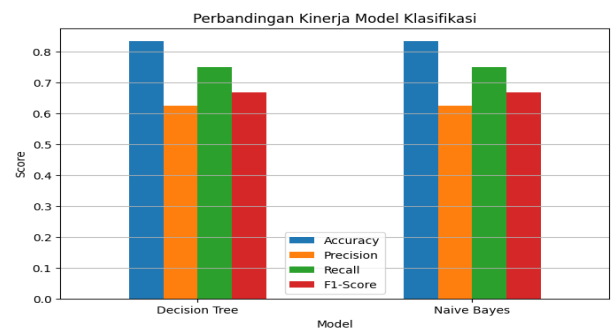
C. *Perbandingan Performa Model Klasifikasi*

Perbandingan performa model dilakukan untuk mengevaluasi efektivitas algoritma Decision Tree dan Naive Bayes dalam mengklasifikasikan kinerja guru berdasarkan dataset yang digunakan. Perbandingan ini didasarkan pada metrik evaluasi utama yang diperoleh dari classification report, yaitu akurasi, presisi, recall, dan F1-score.

TABEL VI
PERBANDINGAN PERFORMA MODEL KLASIFIKASI

Model	Accuracy	Precision	Recall	F1-Score
Decision Tree	0.83	0.62	0.75	0.67
Naive Bayes	0.83	0.62	0.75	0.67

Catatan: Nilai presisi, recall, dan F1-score yang digunakan merupakan macro average untuk merepresentasikan performa model secara seimbang pada seluruh kategori kinerja guru.



Gambar 8. Perbandingan Kinerja Model Klasifikasi

Gambar diatas menampilkan perbandingan kinerja model Decision Tree dan Naive Bayes berdasarkan metrik evaluasi utama, yaitu akurasi, presisi, recall, dan F1-score. Visualisasi ini bertujuan untuk memberikan gambaran komparatif secara intuitif mengenai performa kedua model dalam mengklasifikasikan kinerja guru.

Berdasarkan diagram tersebut, terlihat bahwa kedua model menghasilkan nilai akurasi yang sama, yaitu sebesar 0,83,

yang menunjukkan bahwa secara umum kemampuan prediksi kedua model relatif setara. Namun demikian, perbedaan kinerja mulai terlihat ketika ditinjau dari metrik presisi, recall, dan F1-score.

Nilai recall pada kedua model menunjukkan kecenderungan yang lebih tinggi dibandingkan presisi, yang mengindikasikan bahwa model cenderung lebih baik dalam mengenali data aktual dari masing-masing kategori kinerja dibandingkan dengan ketepatan prediksi kelas tersebut. Hal ini relevan dalam konteks evaluasi kinerja guru, di mana kemampuan model untuk mengenali kategori kinerja yang benar menjadi aspek penting.

Nilai F1-score yang relatif seimbang pada kedua model menunjukkan adanya kompromi antara presisi dan recall. Meskipun demikian, Decision Tree menunjukkan stabilitas performa yang lebih baik berdasarkan hasil confusion matrix dan classification report, sehingga lebih sesuai digunakan sebagai model utama dalam penelitian ini.

Secara keseluruhan, visualisasi ini memperkuat hasil analisis kuantitatif yang telah disajikan pada tabel perbandingan model dan mendukung kesimpulan bahwa Decision Tree merupakan algoritma yang lebih interpretatif dan adaptif terhadap karakteristik data kinerja guru.

Perlu diperhatikan bahwa kategori kinerja guru dalam penelitian ini dibentuk berdasarkan interval nilai PAK yang bersifat deterministik. Selain itu, penggunaan data administratif sebagai sumber utama dalam penelitian ini juga berpotensi mengandung bias struktural, karena indikator PAK lebih merepresentasikan aktivitas yang terdokumentasi secara formal dibandingkan kualitas pedagogis guru secara langsung di dalam proses pembelajaran. Oleh karena itu, hasil klasifikasi perlu diinterpretasikan dengan mempertimbangkan keterbatasan karakteristik data yang digunakan. Dengan demikian, performa model yang relatif tinggi dipengaruhi oleh struktur aturan tersebut. Model klasifikasi dalam penelitian ini lebih merepresentasikan kemampuan algoritma dalam mengotomatisasi aturan administratif yang telah ditetapkan, dibandingkan menemukan pola laten baru dalam data. Selain itu, penggunaan validasi silang membantu memastikan bahwa performa model tidak hanya disebabkan oleh pembagian data tertentu, sehingga risiko overfitting dapat diminimalkan.

D. Implikasi Etis dan Transparansi Model

Penerapan machine learning dalam evaluasi kinerja guru juga memiliki implikasi etis yang perlu dipertimbangkan. Penggunaan model klasifikasi sebagai alat bantu pengambilan keputusan harus disertai dengan prinsip transparansi dan akuntabilitas. Model yang digunakan dalam penelitian ini, khususnya Decision Tree, memiliki keunggulan dalam interpretabilitas karena menghasilkan aturan keputusan yang dapat ditelusuri secara jelas. Namun demikian, hasil klasifikasi tidak seharusnya dijadikan sebagai satu-satunya

dasar dalam penilaian kinerja guru, melainkan sebagai alat pendukung (decision support system) yang tetap memerlukan pertimbangan profesional dari pihak manajemen sekolah. Selain itu, risiko kesalahan klasifikasi (misclassification) harus diantisipasi agar tidak menimbulkan dampak negatif terhadap individu yang dievaluasi.

IV. KESIMPULAN

A. Kesimpulan

Penelitian ini bertujuan untuk mengklasifikasikan kinerja guru menggunakan pendekatan machine learning berbasis data administratif dan nilai Penetapan Angka Kredit (PAK). Dataset yang digunakan terdiri dari 30 data guru yang telah melalui tahapan preprocessing, pembagian data secara stratified, serta validasi silang menggunakan 5-fold cross-validation.

Hasil penelitian menunjukkan bahwa algoritma Decision Tree dan Naive Bayes menghasilkan nilai akurasi sebesar 0,83. Decision Tree menunjukkan performa yang lebih stabil serta memiliki keunggulan dalam aspek interpretabilitas model dibandingkan Naive Bayes. Namun demikian, performa model dipengaruhi oleh struktur kategorisasi kinerja yang bersifat deterministik berdasarkan interval nilai PAK.

Dengan demikian, model klasifikasi dalam penelitian ini lebih merepresentasikan kemampuan algoritma dalam mengotomatisasi aturan administratif yang telah ditetapkan, dibandingkan mengidentifikasi pola laten baru dalam data. Meskipun demikian, pendekatan ini menunjukkan potensi dalam mendukung sistem evaluasi kinerja guru yang lebih terstruktur, transparan, dan berbasis data.

B. Keterbatasan Penelitian

Penelitian ini memiliki beberapa keterbatasan. Pertama, jumlah dataset yang digunakan relatif kecil dan hanya berasal dari satu institusi, sehingga kemampuan generalisasi model masih terbatas. Kedua, kategori kinerja guru dibentuk berdasarkan interval nilai PAK yang bersifat deterministik, sehingga model lebih merepresentasikan otomatisasi aturan administratif dibandingkan eksplorasi pola laten baru. Ketiga, atribut yang digunakan masih terbatas pada data administratif dan belum mencakup variabel kualitatif seperti observasi pembelajaran atau evaluasi perilaku profesional guru. Oleh karena itu, penelitian selanjutnya disarankan untuk menggunakan dataset yang lebih besar, multi-institusi, serta mempertimbangkan penambahan variabel yang lebih beragam untuk meningkatkan validitas eksternal model.

C. Saran

Berdasarkan hasil penelitian, terdapat beberapa rekomendasi untuk pengembangan penelitian selanjutnya. Pertama, penelitian dapat dikembangkan dengan menambah jumlah data serta memperluas cakupan institusi agar model memiliki kemampuan generalisasi yang lebih baik dan

validitas eksternal yang lebih kuat. Kedua, penggunaan algoritma klasifikasi lain serta teknik penyeimbangan data (data balancing) dapat dipertimbangkan untuk meningkatkan performa model pada kategori kinerja dengan jumlah data yang terbatas.

Selain itu, penelitian lanjutan dapat mempertimbangkan penambahan variabel non-administratif, seperti hasil observasi pembelajaran atau indikator kualitatif lainnya, untuk memperkaya representasi kinerja guru. Hasil penelitian ini diharapkan dapat menjadi referensi awal dalam pengembangan sistem pendukung keputusan berbasis data untuk evaluasi kinerja guru secara lebih terstruktur, transparan, dan akuntabel.

DAFTAR PUSTAKA

- [1] H. Kusumaningrum, C. Chaerany, T. A. Kholisah, and R. Cahyani, "Journal Education and Government Wiyata Kinerja Guru Sebagai Aspek Strategis Dalam Pengelolaan Sumber Daya Manusia Pendidikan," vol. 2, no. 2, pp. 105–125, 2024.
- [2] S. Ismail et al., "Pengaruh penggunaan sistem pengukuran kinerja guru di SMA negeri se-Kota Gorontalo," 2024.
- [3] J. Chaidir and J. Friadi, "Zona Teknik: Jurnal Ilmiah Evaluasi Kinerja Algoritma Klasifikasi dalam Studi Kasus Prediksi Kelulusan di Universitas XYZ," vol. 19, no. 1, pp. 15–22, 2025.
- [4] D. Arisandi and T. Sutrisno, "Evaluasi Penilaian Kinerja Karyawan Dengan Metode Naïve Bayes," 2024.
- [5] M. Y. T. Sulistyono, E. S. Pane, E. M. Yuniarno, and M. H. Purnono, "Hybrid Significant Stroke Feature: A Novel Stroke Feature Analysis Approach for Stroke Severity Classification of EEG Signals Based on Time Domain, Frequency Domain, and Signal Decomposition Domain," *International Journal of Intelligent Engineering and Systems*, vol. 17, no. 6, pp. 1241–1267, 2024, doi: 10.22266/ijies2024.1231.91.
- [6] L. Nur Halimah, S. Riyadi, A. Fatahillah Jurjani, A. Prayogi, and S. Dwi Laksana, "Implementasi Penggunaan Machine Learning Dalam Pembelajaran: Suatu Telaah Deskriptif," 2025.
- [7] I. Rusydi and S. R. Nur Aisyiyah, "Implementasi Data Mining terhadap Evaluasi Kinerja Guru dalam Mengajar Menggunakan Metode Naive Bayes Classifier," *VISA: Journal of Visions and Ideas*, vol. 4, no. 1, p. 117, 2024.
- [8] M. Habibullah Arief and M. Kholila Fadhil, "Educational Data Mining for Student Academic Performance Analysis," 2024.
- [9] M. Bellaj, A. Ben Dahmane, and L. Sefian, "Educational Data Mining: Employing Machine Learning Techniques and Hyperparameter Optimization to Improve Students' Academic Performance," *International journal of online and biomedical engineering*, vol. 20, no. 3, pp. 55–74, 2024, doi: 10.3991/ijoe.v20i03.46287.
- [10] R. Sharma and P. K. Singh, "A comparative study of Decision Tree and Naive Bayes algorithms for classification performance evaluation," *Expert Systems with Applications*, vol. 213, art. no. 118995, 2023, doi: 10.1016/j.eswa.2022.118995.
- [11] S. Sucipto, D. Dwi Prasetya, and T. Widiyaningtyas, "Educational Data Mining: Multiple Choice Question Classification in Vocational School," *MATRIK: Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer*, vol. 23, no. 2, pp. 379–388, Mar. 2024, doi: 10.30812/matrik.v23i2.3499.
- [12] M. B. Hartanto, T. Destanto, Y. Yuniarthe, and T. Winarko, "Implementation of Data Mining for Classifying Student Graduation Levels Using Naive Bayes, Decision Tree, Random Forest, Support Vector Machines and Neural Networks Methods (Case Study of The Undergraduate Program at Mitra Indonesia University)," vol. 18, no. 1, 2025.
- [13] T. Gori, A. Sunyoto, and H. Al Fatta, "Preprocessing Data dan Klasifikasi untuk Prediksi Kinerja Akademik Siswa," *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 11, no. 1, pp. 215–224, Feb. 2024.
- [14] M. Riyadi Maskur and A. Wibowo, "Taxpayer Awareness Classification Using Decision Tree and Naïve Bayes Methods," 2024.
- [15] M. A. Hasanah, S. Soim, and A. S. Handayani, "Implementasi CRISP-DM Model Menggunakan Metode Decision Tree dengan Algoritma CART untuk Prediksi Curah Hujan Berpotensi Banjir," 2021.
- [16] A. Ahmed et al., "Students' performance prediction employing Decision Tree," *CTU Journal of Innovation and Sustainable Development*, vol. 16, no. Special Issue, pp. 42–51, Oct. 2024, doi: 10.22144/ctujoisd.2024.321.
- [17] D. Kurniasari, R. N. Hidayah, N. Notiragayu, W. Warsono, and R. K. Nisa, "Classification Models For Academic Performance: A Comparative Study Of Naïve Bayes And Random Forest Algorithms In Analyzing University Of Lampung Student Grades," *Jurnal Teknik Informatika (Jutif)*, vol. 5, no. 5, pp. 1267–1276, Oct. 2024, doi: 10.52436/1.jutif.2024.5.5.2066.
- [18] H. Amalia, A. Puspita, A. F. Lestari, and F. Friyadie, "Application Of Decision Tree And Naive Bayes On Student Performance Dataset," *Jurnal Pilar Nusa Mandiri*, vol. 18, no. 1, pp. 53–58, Mar. 2022, doi: 10.33480/pilar.v18i1.2714.
- [19] A. Rahman, "Klasifikasi Performa Akademik Siswa Menggunakan Metode Decision Tree dan Naive Bayes," *Jurnal SAINTEKOM*, vol. 13, no. 1, pp. 22–31, Mar. 2023, doi: 10.33020/saintekom.v13i1.349.
- [20] P. Wahyuni and M. A. Romli, "Comparison of Naïve Bayes Classifier and Decision Tree Algorithms for Sentiment Analysis on the House of Representatives' Right of Inquiry on Twitter," 2024.
- [21] A.-Husaini, I. Hariyanti, and A. R. Raharja, "Perbandingan Algoritma Decision Tree dan Naive Bayes dalam Klasifikasi Data Pengaruh Media Sosial dan Jam Tidur Terhadap Prestasi Akademik Siswa," *Technologia : Jurnal Ilmiah*, vol. 15, no. 2, p. 332, Apr. 2024, doi: 10.31602/tji.v15i2.14381.