Indonesian Journal of Multidisciplinary on Social and Technology
Homepage: https://journal.ilmudata.co.id/index.php/ijmst
Vol. 4 No. 3 (2026) pp: 10886-10894
P-ISSN: 2986-6790, e-ISSN: 2986-6782

Analisis Penilaian Portofolio Potensi Siswa Menggunakan Metode K-Nearst Neighbor

Firgilius Bili[1], Friden Elefri Neno[2], Sihang Gregorius Bali Mema[3]
[1]Teknik Informatika, Teknik, Universitas Stella Maris Sumba
[2]Teknik Informatika, Teknik, Universitas Stella Maris Sumba
[3]Administrasi Rumah Sakit, Fakultas Kesehatan, Universitas Stella Maris Sumba
[1]bilifirgo@gmail.com, [2]nenofriden.e@gmail.com, [3]sihanggregoriusbalimema @gmail.com

Abstrak

Penilaian potensi siswa merupakan aspek penting dalam pendidikan untuk mengidentifikasi kemampuan akademik dan nonakademik secara sistematis. Namun, penilaian portofolio yang melibatkan berbagai indikator memerlukan pendekatan analitis untuk membantu proses klasifikasi potensi siswa secara terukur. Penelitian ini bertujuan menganalisis penerapan algoritma K-Nearest Neighbor (KNN) dalam mengklasifikasikan potensi siswa berdasarkan atribut penilaian portofolio. Dataset yang digunakan terdiri atas 100 data simulasi dengan lima atribut prediktor, yaitu nilai akademik, prestasi, keterampilan, kreativitas, dan kedisiplinan. Variabel target dikelompokkan menjadi tiga kategori, yaitu tinggi, sedang, dan rendah. Tahapan penelitian meliputi pemeriksaan kelengkapan data, preprocessing, normalisasi menggunakan Min-Max Scaling, pembagian data pelatihan dan pengujian dengan rasio 80:20, perhitungan jarak Euclidean, klasifikasi KNN, serta evaluasi model. Pengujian dilakukan menggunakan metrik accuracy, precision, recall, Macro-F1-score, dan confusion matrix. Hasil pengujian menunjukkan bahwa model memperoleh accuracy sebesar 95%, Macro-Precision sebesar 96,67%, Macro-Recall sebesar 95,24%, dan Macro-F1-score sebesar 95,68%. Berdasarkan confusion matrix, sebanyak 19 dari 20 data pengujian diklasifikasikan dengan benar, sedangkan satu data berkategori tinggi diprediksi sebagai sedang. Hasil ini menunjukkan bahwa KNN mampu mempelajari pola klasifikasi pada dataset simulasi yang digunakan. Namun, karena label target dibentuk berdasarkan rata-rata atribut prediktor yang sama, hasil tersebut belum membuktikan validitas identifikasi potensi siswa dalam kondisi nyata. Penelitian selanjutnya perlu menggunakan data portofolio aktual, rubrik penilaian tervalidasi, dan pembandingan algoritma untuk menguji generalisasi

Kata kunci: K-Nearest Neighbor, Penilaian Portofolio, Potensi Siswa, Klasifikasi, Machine Learning.

Abstract

Assessing student potential is essential in education to systematically identify academic and non-academic abilities. However, portfolio assessment involving multiple indicators requires an analytical approach to support measurable classification of student potential. This study aims to analyze the application of the K-Nearest Neighbor (KNN) algorithm in classifying student potential based on portfolio assessment attributes. The dataset consisted of 100 simulated records with five predictor attributes: academic scores, achievements, skills, creativity, and discipline. The target variable was classified into three categories: high, moderate, and low. The research stages included data completeness checking, preprocessing, Min-Max Scaling normalization, an 80:20 training-testing split, Euclidean distance calculation, KNN classification, and model evaluation. Performance was evaluated using accuracy, precision, recall, Macro-F1-score, and a confusion matrix. The results demonstrated an accuracy of 95%, Macro-Precision of 96.67%, Macro-Recall of 95.24%, and Macro-F1-score of 95.68%. The confusion matrix indicated that 19 of 20 test samples were correctly classified, while one high-category sample was misclassified as moderate. These findings demonstrate that KNN effectively learned classification patterns within the simulated dataset. However, because the target labels were generated from the average of the same predictor attributes, these results do not establish the validity of student potential identification in real-world settings. Future research should incorporate actual portfolio data, validated assessment rubrics, and comparative algorithm evaluations to examine model generalizability.

Keywords: K-Nearest Neighbor, Portfolio Assessment, Student Potential, Classification, Machine Learning

1. Pendahuluan

Penilaian potensi siswa merupakan salah satu aspek penting dalam proses pendidikan karena dapat membantu pendidik memahami kemampuan akademik, keterampilan, kreativitas, prestasi, dan perkembangan kompetensi siswa secara menyeluruh. Penilaian yang hanya berorientasi pada nilai akademik berpotensi belum

menggambarkan seluruh kemampuan siswa, sehingga diperlukan pendekatan evaluasi yang mempertimbangkan berbagai bukti hasil belajar. Penilaian portofolio dapat digunakan untuk mengumpulkan dan mengevaluasi karya, proyek, prestasi, serta aktivitas siswa selama periode pembelajaran. Pendekatan tersebut memungkinkan pendidik memperoleh informasi mengenai perkembangan kompetensi siswa berdasarkan bukti pembelajaran yang terdokumentasi (Burner, 2014; Lam, 2016; Schellekens et al., 2021).

Meskipun demikian, penilaian portofolio menghadapi tantangan ketika jumlah siswa dan indikator penilaian semakin beragam. Evaluasi secara manual dapat membutuhkan waktu yang relatif lama serta berpotensi menghasilkan inkonsistensi apabila kriteria penilaian tidak dirumuskan secara jelas. Selain itu, pengolahan data dari berbagai indikator akademik dan nonakademik membutuhkan pendekatan analitis yang mampu mengidentifikasi pola dan hubungan ant atribut secara sistematis. Penelitian dalam bidang Educational Data Mining menunjukkan bahwa data pendidikan dapat dianalisis menggunakan teknik komputasi untuk mendukung pemahaman terhadap kinerja dan perkembangan siswa (Shahiri et al., 2015; Zhang et al., 2021).

Perkembangan machine learning memberikan peluang untuk mengembangkan sistem klasifikasi yang dapat memanfaatkan data historis siswa sebagai dasar untuk mengidentifikasi pola tertentu. Berbagai algoritma klasifikasi telah digunakan untuk menganalisis kinerja akademik dan memprediksi hasil belajar, termasuk metode berbasis tetangga terdekat. Pendekatan Educational Data Mining juga memungkinkan proses evaluasi dilakukan berdasarkan atribut yang terukur dan hasilnya dievaluasi menggunakan metrik klasifikasi. Namun, kinerja model sangat bergantung pada kualitas data, relevansi fitur, serta ketepatan prosedur evaluasi yang diterapkan (Bin Roslan & Chen, 2022; Chaka, 2022; Zhang et al., 2021).

Salah satu algoritma yang dapat diterapkan adalah K-Nearest Neighbor (KNN), yaitu metode klasifikasi yang menentukan kelas suatu data berdasarkan mayoritas kelas dari sejumlah tetangga terdekat dalam ruang fitur. Prinsip KNN didasarkan pada pengukuran kemiripan atau jarak antara data yang akan diklasifikasikan dan data pelatihan. Cover dan Hart (1967) membahas landasan teoretis klasifikasi tetangga terdekat, yang kemudian menjadi dasar bagi berbagai penerapan algoritma berbasis jarak. Dalam penelitian ini, KNN dipertimbangkan untuk mengklasifikasikan potensi siswa berdasarkan kombinasi atribut nilai akademik, prestasi, keterampilan, kreativitas, dan kedisiplinan.

Penerapan KNN pada data portofolio membutuhkan perhatian terhadap perbedaan skala antaratribut, pemilihan parameter \(k\), distribusi kelas, dan kualitas label target. Atribut yang mempunyai rentang nilai lebih besar dapat mendominasi perhitungan jarak apabila data tidak dinormalisasi. Oleh karena itu, normalisasi fitur, pembagian data pelatihan dan pengujian secara terstratifikasi, serta pemilihan parameter model harus dirancang secara tepat. Evaluasi juga perlu menggunakan metrik yang sesuai untuk klasifikasi multikelas, seperti accuracy, precision, recall, Macro-F1-score, dan confusion matrix, agar performa model tidak hanya dinilai berdasarkan satu ukuran saja (Cover & Hart, 1967; Shahiri et al., 2015; Zhang et al., 2021).

Berdasarkan kajian tersebut, penelitian mengenai klasifikasi kinerja akademik menggunakan Educational Data Mining telah berkembang, tetapi penerapan KNN secara khusus untuk mengklasifikasikan potensi siswa berdasarkan kombinasi indikator penilaian portofolio masih perlu dikaji melalui penelitian empiris yang terstruktur. Kesenjangan penelitian yang hendak ditelaah adalah bagaimana kombinasi atribut akademik dan nonakademik dapat digunakan untuk mengklasifikasikan siswa ke dalam kategori potensi tinggi, sedang, dan rendah, serta bagaimana pemilihan parameter dan prosedur prapemrosesan memengaruhi kinerja klasifikasi. Kesenjangan ini perlu dikonfirmasi melalui perbandingan sistematis dengan penelitian terdahulu yang memiliki objek, variabel, dan metode serupa (Bin Roslan & Chen, 2022; Chaka, 2022; Zhang et al., 2021).

Penelitian ini bertujuan menganalisis penerapan algoritma K-Nearest Neighbor untuk mengklasifikasikan potensi siswa berdasarkan data penilaian portofolio. Tahapan penelitian meliputi pengumpulan data, prapemrosesan, normalisasi atribut, pembagian data pelatihan dan pengujian, penentuan parameter \(k\), serta evaluasi menggunakan accuracy, precision, recall, Macro-F1-score, dan confusion matrix. Kontribusi yang diharapkan adalah menyediakan evaluasi empiris mengenai kelayakan KNN untuk klasifikasi potensi siswa serta memberikan dasar analitis bagi pengembangan sistem pendukung evaluasi pendidikan. Sistem tersebut tetap diposisikan sebagai alat bantu bagi pendidik, bukan sebagai pengganti penilaian profesional terhadap kemampuan dan kebutuhan belajar siswa (Schellekens et al., 2021; Zhang et al., 2021).

DOI: https://doi.org/10.69693/ijmst.v4i3.14473
Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)

2. Metode Penelitian

Penerapan algoritma K-Nearest Neighbor (KNN) dalam mengklasifikasikan potensi siswa berdasarkan data penilaian portofolio. Penelitian dilakukan melalui beberapa tahapan, yaitu pengumpulan data, prapemrosesan, penentuan atribut, normalisasi data, pembagian dataset, pemodelan KNN, dan evaluasi kinerja klasifikasi. bertujuan mengukur kemampuan algoritma dalam mengklasifikasikan siswa ke dalam tiga kategori potensi, yaitu tinggi, sedang, dan rendah, berdasarkan atribut penilaian yang telah ditentukan. (Al-Din, M. S. N., et al Abdulqader, H. A.2024)



Gambar 1. Tahapan-Tahapan Penelitian

2.1 Pengumpulan Data Portofolio

Pengumpulan data dalam penelitian ini dilakukan untuk memperoleh atribut penilaian portofolio yang digunakan dalam proses klasifikasi potensi siswa menggunakan algoritma K-Nearest Neighbor (KNN). Data portofolio mencakup lima atribut utama, yaitu nilai akademik, prestasi, keterampilan, kreativitas, dan kedisiplinan. Pemilihan atribut tersebut didasarkan pada pertimbangan bahwa potensi siswa perlu dianalisis melalui kombinasi indikator yang merepresentasikan kemampuan akademik dan nonakademik. Setiap atribut direpresentasikan dalam bentuk skor numerik pada rentang 0–100 untuk memudahkan proses prapemrosesan, normalisasi, dan perhitungan jarak pada algoritma KNN (Al-Shanableh, N,et al,2026)

2.2 Preprocesing dan Normalisasi Data

Tahap data processing dilakukan melalui pemeriksaan kelengkapan data, tipe data, rentang nilai, dan duplikasi pada lima atribut portofolio, yaitu nilai akademik, prestasi, keterampilan, kreativitas, dan kedisiplinan. Selanjutnya, atribut prediktor dipisahkan dari label target potensi siswa yang terdiri atas kelas tinggi, sedang, dan rendah. Normalisasi menggunakan metode Min-Max Scaling dengan rentang 0–1 untuk menyetarakan skala atribut sebelum perhitungan jarak Euclidean pada algoritma K-Nearest Neighbor (KNN). Parameter normalisasi dihitung berdasarkan data pelatihan untuk menghindari data leakage. Dataset kemudian dibagi menjadi 80% data pelatihan dan 20% data pengujian secara terstratifikasi untuk mengevaluasi kinerja model secara objektif Alwarthan, S. A,et al,2022

2.3 Pembagian Data Training dan Data Testing

Penelitian ini membahas penerapan algoritma K-Nearest Neighbor (KNN) untuk mengklasifikasikan potensi siswa berdasarkan penilaian portofolio yang mencakup nilai akademik, prestasi, keterampilan, kreativitas, dan kedisiplinan. Tahapan penelitian meliputi data processing, normalisasi menggunakan Min-Max Scaling,

DOI: https://doi.org/10.69693/ijmst.v4i3.14473
Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)

pembagian data training sebesar 80% dan testing sebesar 20%, serta evaluasi model menggunakan accuracy, precision, recall, dan Macro-F1-score. Pendekatan ini diharapkan dapat mendukung analisis potensi siswa secara lebih sistematis dan terukur. Namun, karena dataset yang digunakan masih berupa data simulasi sebanyak 100 baris, efektivitas model belum dapat disimpulkan untuk kondisi pendidikan nyata. Penelitian selanjutnya perlu menggunakan data portofolio siswa aktual yang telah dianonimkan, label yang divalidasi oleh pendidik, dan perbandingan dengan algoritma klasifikasi lainnya untuk memperoleh hasil yang lebih objektif dan dapat dipertanggungjawabkan secara ilmiah. (Apeadido, S,at al,2024)

2.4 Perhitungan Jarak dengan KNN

Perhitungan jarak pada algoritma K-Nearest Neighbor (KNN) dilakukan untuk mengukur tingkat kemiripan antara data siswa yang diuji dan data siswa pada training. Penelitian ini menggunakan jarak Euclidean berdasarkan lima atribut portofolio, yaitu nilai akademik, prestasi, keterampilan, kreativitas, dan kedisiplinan. Rumus jarak Euclidean dinyatakan sebagai berikut:

$$d(x,y) = \sqrt{\sum (xi - yi)^2}$$

Keterangan: d(x, y) adalah jarak antara dua data; x_i adalah nilai atribut data yang diuji; y_i adalah nilai atribut data training; dan n adalah jumlah atribut

Dengan x_i merupakan nilai atribut siswa yang diuji, y_i merupakan nilai atribut siswa pada data *training*, dan *n* adalah jumlah atribut yang digunakan. Setelah jarak setiap data dihitung, algoritma mengurutkan hasilnya dari jarak terkecil hingga terbesar, kemudian memilih sebanyak *k* tetangga terdekat. Kategori potensi siswa ditentukan berdasarkan kelas mayoritas dari tetangga tersebut, yaitu tinggi, sedang, atau rendah. Nilai *k* dapat diuji menggunakan beberapa konfigurasi, seperti $k = 3$ dan $k = 5$ dan $k = 7$, untuk menentukan parameter yang memberikan kinerja klasifikasi terbaik berdasarkan validasi pada data *training*. Selanjutnya, konfigurasi terpilih dievaluasi menggunakan data *testing* yang tidak digunakan dalam pemilihan parameter (Bin Roslan, M. H,et al,2024),(Chaka, C. 2022).

2.5 Klasifikasi Potensi Siswa

Klasifikasi potensi siswa dilakukan menggunakan algoritma K-Nearest Neighbor (KNN) dengan menentukan kategori berdasarkan mayoritas kelas dari \(k\) tetangga terdekat. Data siswa diklasifikasikan ke dalam tiga kategori, yaitu Tinggi, Sedang, dan Rendah, berdasarkan kemiripan lima atribut portofolio: nilai akademik, prestasi, keterampilan, kreativitas, dan kedisiplinan. Setelah jarak Euclidean dihitung dan diurutkan dari nilai terkecil, algoritma memilih sejumlah \(k\) tetangga terdekat. Kelas dengan jumlah anggota terbanyak menjadi hasil prediksi potensi siswa. Apabila terjadi jumlah suara yang sama, aturan pemecah seri perlu ditetapkan secara konsisten. Pemilihan nilai \(k\) dilakukan melalui validasi pada data pelatihan, sedangkan kinerja klasifikasi akhir dievaluasi menggunakan data pengujian dengan metrik accuracy, precision, recall, dan Macro-F1-score. (Guanin-Fajardo, J. H,2024), Indahyanti, U., Azizah,at al,2024)

2.6 Evaluasi Model dan Potensi Hasil

Evaluasi model K-Nearest Neighbor (KNN) dilakukan menggunakan metrik accuracy, precision, recall, Macro-F1-score, dan confusion matrix untuk mengukur kemampuan klasifikasi potensi siswa ke dalam kategori tinggi, sedang, dan rendah. Hasil klasifikasi digunakan untuk menganalisis potensi siswa berdasarkan atribut akademik dan nonakademik sebagai bahan pendukung evaluasi serta perencanaan pembelajaran. Penelitian ini diharapkan menghasilkan model klasifikasi yang sistematis dan terukur. Namun, nilai evaluasi harus berdasarkan hasil pengujian aktual, sedangkan validitas penerapan pada siswa nyata memerlukan pengujian menggunakan data empiris dan label yang tervalidasi (Jayasree, R,at al,2023)

DOI: https://doi.org/10.69693/ijmst.v4i3.14473
Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)

3. Hasil dan Pembahasan

3.1 Pengumpulan Data

Dataset penilaian portofolio siswa yang digunakan sebagai masukan dalam klasifikasi potensi siswa menggunakan algoritma K-Nearest Neighbor (KNN). Dataset terdiri atas atribut identitas siswa, nilai akademik, prestasi, keterampilan, kreativitas, kedisiplinan, rata-rata portofolio, dan label potensi siswa. Setiap atribut numerik memiliki rentang skor 0–100, sedangkan label target dikategorikan menjadi tiga kelas, yaitu Tinggi, Sedang, dan Rendah. Sebagai contoh, siswa S001 memiliki nilai rata-rata portofolio sebesar 80,4 dan dikategorikan sebagai Tinggi, sedangkan siswa S002 memiliki rata-rata 62,0 dan dikategorikan sebagai Sedang. Sementara itu, siswa S005 memiliki rata-rata 35,0 dan termasuk kategori Rendah. Data tersebut selanjutnya diproses melalui tahapan preprocessing, normalisasi menggunakan Min-Max Scaling, dan pembagian data menjadi training dan testing. Dalam penerapan KNN, lima atribut penilaian nilai akademik, prestasi, keterampilan, kreativitas, dan kedisiplinan—digunakan untuk menghitung jarak Euclidean antara data siswa yang diuji dan data pelatihan. Kategori potensi ditentukan berdasarkan mayoritas kelas dari sejumlah tetangga terdekat. Kolom ID_Siswa tidak digunakan sebagai fitur prediktor, sedangkan Rata_Rata_Portofolio perlu dikeluarkan dari fitur apabila label potensi dibentuk berdasarkan nilai rata-rata tersebut, guna menghindari data leakage

	ID_Siswa	Nilai_Akademik	Prestasi	Keterampilan	Kreativitas	Kedisiplinan	Rata_Rata_Portofolio	Potensi_Siswa
0	S001	73	77	91	71	90	80.4	Tinggi
1	S002	69	69	42	57	73	62.0	Sedang
2	S003	79	80	88	89	76	82.4	Tinggi
3	S004	90	90	78	78	71	81.4	Tinggi
4	S005	39	14	46	47	29	35.0	Rendah

Gambar 2. Pengumpulan Data

3.2 Preprocessing dan normalisasi Data

Jumlah nilai kosong per kolom:

	Jumlah_Kosong
ID_Siswa	0
Nilai_Akademik	0
Prestasi	0
Keterampilan	0
Kreativitas	0
Kedisiplinan	0
Rata_Rata_Portofolio	0
Potensi_Siswa	0

Jumlah baris awal: 100
Jumlah baris setelah pemeriksaan: 100

Gambar 3. Preprocessing dan normalisasi Data

Gambar 3 menunjukkan hasil pemeriksaan nilai kosong (missing values) pada dataset penilaian portofolio siswa. Hasil pemeriksaan menunjukkan bahwa seluruh kolom, meliputi ID_Siswa, Nilai_Akademik, Prestasi, Keterampilan, Kreativitas, Kedisiplinan, Rata_Rata_Portofolio, dan Potensi_Siswa, memiliki jumlah nilai kosong sebesar nol. Jumlah data sebelum dan sesudah proses pemeriksaan tetap sebanyak 100 baris. Hasil tersebut menunjukkan bahwa seluruh observasi memiliki nilai pada setiap kolom yang diperiksa sehingga tidak diperlukan penanganan nilai kosong. Pemeriksaan ini menjadi tahap awal untuk memastikan kelengkapan data sebelum proses klasifikasi menggunakan algoritma K-Nearest Neighbor (KNN).

DOI: https://doi.org/10.69693/ijmst.v4i3.14473
Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)

Distribusi label target:

Potensi_Siswa	Jumlah
Sedang	46
Tinggi	35
Rendah	19

Fitur yang digunakan: ['Nilai Akademik', 'Prestasi', 'Keterampilan', 'Kreativitas', 'Kedisiplin

Gambar 4.Distribusi Laber Target

Gambar 4. menunjukkan distribusi label target Potensi_Siswa dan atribut yang digunakan dalam pemodelan. Dari total 100 data, sebanyak 46 siswa (46%) termasuk kategori Sedang, 35 siswa (35%) termasuk kategori Tinggi, dan 19 siswa (19%) termasuk kategori Rendah. Distribusi tersebut menunjukkan bahwa kelas Sedang memiliki jumlah observasi terbesar, sedangkan kelas Rendah memiliki jumlah observasi terkecil. Kondisi ini menunjukkan adanya ketidakseimbangan kelas yang perlu diperhatikan ketika mengevaluasi kinerja model. Lima atribut yang digunakan sebagai fitur prediktor adalah Nilai_Akademik, Prestasi, Keterampilan, Kreativitas, dan Kedisiplinan, sedangkan Potensi_Siswa berperan sebagai variabel target. Kolom ID_Siswa tidak digunakan sebagai fitur karena hanya berfungsi sebagai identitas observasi. Kolom Rata_Rata_Portofolio

3.3 Pembagian data training dan testing (80:20, stratified)

Jumlah data testing : 20

Proporsi kelas pada training:

Potensi_Siswa	Persen
Sedang	46.25
Tinggi	35.00
Rendah	18.75

Proporsi kelas pada testing:

Potensi_Siswa	Persen
Sedang	45.0
Tinggi	35.0
Rendah	20.0

Gambar 5. Pembagian data training dan testing (80:20, stratified)

Pembagian data 80:20 telah menghasilkan 80 data training dan 20 data testing, dengan proporsi kelas yang relatif konsisten pada kedua kelompok. Dataset selanjutnya dapat digunakan untuk proses normalisasi, pemilihan parameter \(k\), klasifikasi potensi siswa, dan evaluasi menggunakan accuracy, precision, recall, Macro-F1-score, serta confusion matrix.

DOI: https://doi.org/10.69693/ijmst.v4i3.14473
Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)

3.4 Perhitungan jarak KNN

	Nilai_Akademik	Prestasi	Keterampilan	Kreativitas	Kedisiplinan	Label_Aktual	Label_Prediksi	Benar
6	66	60	68	69	98	Sedang	Sedang	True
10	87	92	92	65	70	Tinggi	Tinggi	True
69	69	65	76	56	75	Sedang	Sedang	True
26	91	100	61	85	93	Tinggi	Tinggi	True
56	69	69	69	77	65	Sedang	Sedang	True
53	64	51	74	87	90	Sedang	Sedang	True
32	70	64	57	60	66	Sedang	Sedang	True
65	75	92	94	75	90	Tinggi	Tinggi	True
94	41	40	57	42	47	Rendah	Rendah	True
70	50	40	55	55	49	Rendah	Rendah	True
55	78	68	65	62	66	Sedang	Sedang	True
46	78	68	75	93	86	Tinggi	Tinggi	True

Gmbar 6. Perhitungan jarak KNN

Berdasarkan gambar, algoritma K-Nearest Neighbor (KNN) berhasil mengklasifikasikan data pengujian ke dalam kategori potensi Tinggi, Sedang, dan Rendah. Seluruh baris yang terlihat menunjukkan kesesuaian antara Label_Aktual dan Label_Prediksi, yang ditandai dengan nilai True pada kolom Benar. Hasil ini menunjukkan bahwa prediksi model sesuai dengan label aktual pada sampel yang ditampilkan. Namun, karena gambar hanya memperlihatkan sebagian data pengujian, nilai akurasi keseluruhan belum dapat ditentukan. Evaluasi akhir perlu menggunakan seluruh data testing melalui metrik accuracy, precision, recall, Macro-F1-score, dan confusion matrix.

3.5 Evaluasi Model

```
Accuracy       : 0.9500
Macro-Precision: 0.9667
Macro-Recall   : 0.9524
Macro-F1       : 0.9568

Classification report:
              precision    recall  f1-score   support

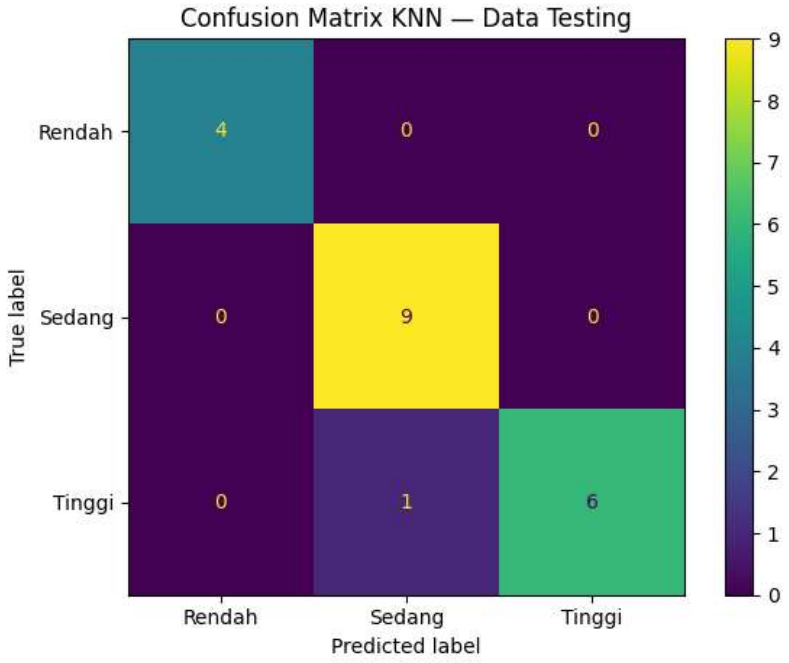
      Rendah       1.00      1.00      1.00         4
      Sedang       0.90      1.00      0.95         9
      Tinggi       1.00      0.86      0.92         7

    accuracy                           0.95        20
   macro avg       0.97      0.95      0.96        20
weighted avg       0.96      0.95      0.95        20
```

Gambar 7. Evaluasi Model K-Nearest Neighbor (KNN)

Hasil evaluasi menunjukkan bahwa algoritma KNN memperoleh accuracy sebesar 95,00%, Macro-Precision sebesar 96,67%, Macro-Recall sebesar 95,24%, dan Macro-F1-score sebesar 95,68% pada 20 data pengujian. Berdasarkan laporan klasifikasi, kategori Rendah memperoleh precision, recall, dan F1-score sebesar 1,00. Kategori Sedang menghasilkan precision sebesar 0,90, recall sebesar 1,00, dan F1-score sebesar 0,95, sedangkan kategori Tinggi memperoleh precision sebesar 1,00, recall sebesar 0,86, dan F1-score sebesar 0,92. Hasil ini menunjukkan bahwa model memiliki kinerja klasifikasi yang tinggi pada dataset pengujian, meskipun masih terdapat satu kesalahan klasifikasi yang perlu dianalisis lebih lanjut.

DOI: https://doi.org/10.69693/ijmst.v4i3.14473
Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)



Gambar 8.Confusion Matrix KNN

Hasil evaluasi menggunakan confusion matrix menunjukkan bahwa algoritma KNN berhasil mengklasifikasikan 19 dari 20 data pengujian dengan benar, menghasilkan akurasi sebesar 95%. Pada kategori Rendah, seluruh empat data berhasil diklasifikasikan secara tepat, sedangkan sembilan data pada kategori Sedang juga teridentifikasi dengan benar. Pada kategori Tinggi, enam dari tujuh data diklasifikasikan dengan benar dan satu data salah diklasifikasikan sebagai Sedang. Hasil ini menunjukkan bahwa model memiliki kinerja klasifikasi yang baik pada dataset pengujian, dengan kesalahan yang terbatas pada perbedaan antara kategori Tinggi dan Sedang

4. Kesimpulan

Berdasarkan hasil penelitian, algoritma K-Nearest Neighbor (KNN) dapat diterapkan untuk mengklasifikasikan potensi siswa berdasarkan lima atribut penilaian portofolio, yaitu nilai akademik, prestasi, keterampilan, kreativitas, dan kedisiplinan. Pengujian menggunakan 100 data simulasi dengan pembagian 80% data training dan 20% data testing menghasilkan nilai accuracy sebesar 95%, Macro-Precision sebesar 96,67%, Macro-Recall sebesar 95,24%, dan Macro-F1-score sebesar 95,68%. Berdasarkan confusion matrix, model berhasil mengklasifikasikan 19 dari 20 data pengujian dengan benar, dengan satu kesalahan klasifikasi pada kategori tinggi yang diprediksi sebagai sedang. Hasil tersebut menunjukkan bahwa KNN mampu mempelajari pola kategori potensi pada dataset simulasi yang digunakan. Namun, karena label target dibentuk berdasarkan rata-rata atribut yang sama dengan fitur masukan, hasil evaluasi belum membuktikan kemampuan model dalam mengidentifikasi potensi siswa secara independen. Penelitian selanjutnya perlu menggunakan data portofolio siswa nyata yang telah dianonimkan, label yang divalidasi oleh pendidik, serta perbandingan dengan algoritma klasifikasi lain untuk menguji generalisasi dan meningkatkan validitas hasil penelitian.

Reference

1. Bin Roslan, M. H., & Chen, C. J. (2022). Educational data mining for student performance prediction: A systematic literature review (2015–2021). International Journal of Emerging Technologies in Learning, 17(5), 147–179.
2. Burner, T. (2012). The potential formative benefits of portfolio assessment in second and foreign language writing contexts: A review of the literature. Studies in Educational Evaluation, 43, 139–149.
3. Chaka, C. (2022). Educational data mining, student academic performance prediction, prediction methods, algorithms and tools: An overview of reviews. Journal of E-Learning and Knowledge Society, 18(2), 58–69.
4. Cover, T. M., & Hart, P. E. (1967). Nearest neighbor pattern classification. IEEE Transactions on Information Theory, 13(1), 21–27
5. Lam, R. (2016). Assessment as learning: Examining a cycle of teaching, learning, and assessment of writing in the portfolio-based classroom. Studies in Higher Education, 41(11), 1900–1917.

DOI: https://doi.org/10.69693/ijmst.v4i3.14473
Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)

6. Schellekens, L. H., Bok, H. G. J., de Jong, L. H., van der Schaaf, M. F., Kremer, W. D. J., & van der Vleuten, C. P. M. (2021). A scoping review on the notions of assessment as learning (AaL), assessment for learning (AfL), and assessment of learning (AoL). Studies in Educational Evaluation, 71, Article 101094.
7. Shahiri, A. M., Husain, W., & Rashid, N. A. (2015). A review on predicting student's performance using data mining techniques. Procedia Computer Science, 72, 414–422.
8. Zhang, Y., Yun, Y., An, R., Cui, J., Dai, H., & Shang, X. (2021). Educational data mining techniques for student performance prediction: Method review and comparison analysis. Frontiers in Psychology, 12, Article 698490.
9. Al-Din, M. S. N., & Al Abdulqader, H. A. (2024). Students' academic performance prediction using educational data mining and machine learning: A systematic review. International Journal of Research and Innovation in Social Science.
10. Al-Shanableh, N., Alzyoud, M. S., Khalil, A., Benlamine, M., Kraidia, I., Damrah, S., & Tabakhi, A. M. (2026). Forecasting students' academic performance in educational data using machine learning techniques. International Journal of Information and Communication Technology Education, 22
11. Alwarthan, S. A., Aslam, N., & Khan, I. U. (2022). Predicting student academic performance at higher education using data mining: A systematic review. Applied Computational Intelligence and Soft Computing, 2022
12. Apeadido, S., Dzomeku, D., & Nanor, E. (2024). Navigating the portfolio landscape: Examining lecturers' and students' perspectives on portfolio assessment practices. Asian Journal of Education and Social Studies, 50(5), 563–582.
13. Bin Roslan, M. H., & Chen, C. J. (2022). Educational data mining for student performance prediction: A systematic literature review (2015–2021). International Journal of Emerging Technologies in Learning, 17(5), 147–179.
14. Chaka, C. (2022). Educational data mining, student academic performance prediction, prediction methods, algorithms and tools: An overview of reviews. Journal of E-Learning and Knowledge Society, 18(2), 58–69.
15. Guanin-Fajardo, J. H., Guaña-Moya, J., & Casillas, J. (2024). Predicting academic success of college students using machine learning techniques. Data, 9(4), Article 60.
16. Indahyanti, U., Azizah, N. L., & Setiawan, H. (2022). Educational data mining on student academic performance prediction: A survey. Procedia of Social Sciences and Humanities
17. Jayasree, R., & Selvakumari, S. (2023). Design of a prediction model to predict students' performance using educational data mining and machine learning. Engineering Proceedings, 59

DOI: https://doi.org/10.69693/ijmst.v4i3.14473
Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)