

IMPLEMENTASI ALGORITMA K-NEAREST NEIGHBOR UNTUK OPTIMASI PEMBERIAN REWARD SISWA SMA

Agus Riyanto¹, Nurchim², Intan Oktaviani³

^{1,2,3} Jurusan Sistem Informasi Fakultas Ilmu Komputer Universitas Duta Bangsa
Jl. Bhayangkara No.55 – Serengan, Kota Surakarta

¹210101093@mhs.udb.ac.id

²nurchim@udb.ac.id

³intan_oktaviani@udb.ac.id

Abstract

Providing rewards to students is a common strategy to enhance learning motivation and participation in school. However, conventional reward systems often face challenges related to objectivity and the comprehensiveness of assessment criteria. This study aims to apply the K-Nearest Neighbor (KNN) algorithm as an optimization tool for determining reward recipients among high school students. The dataset includes average report card scores, moral values, parental income, number of siblings, and non-academic activity scores. The KNN method was selected for its effectiveness in classifying data based on similarity. The research process involved collecting historical student data, normalizing the data, building the KNN model, and evaluating its performance. The results show that the KNN model achieved an accuracy rate of 91% in classifying students into appropriate reward categories. This indicates that the model can reliably support decision-making in the reward distribution process. In conclusion, applying the KNN algorithm offers a more structured, objective, and transparent approach to student reward allocation, which can enhance the effectiveness of reward programs and promote student development.

Keywords: K-Nearest Neighbor, Reward, Classification, Objectivity, Optimization

I. PENDAHULUAN

Pendidikan mempunyai peranan fundamental dalam membentuk generasi masa depan sebuah bangsa. Dalam konteks Pendidikan formal di sekolah, pemberian penghargaan atau reward merupakan salah satu instrumen penting untuk memotivasi siswa agar terus berprestasi dan mengembangkan potensi diri. *Reward* tidak hanya berfungsi sebagai bentuk apresiasi terhadap pencapaian akademik, tetapi juga terhadap berbagai aspek pengembangan diri yang lain, seperti partisipasi dalam kegiatan non-akademik, pengembangan karakter serta kondisi sosial. Namun sistem pemberian reward yang berlaku umum di sekolah saat ini seringkali masih bersifat konvensional dan memiliki beberapa keterbatasan. Tantangan utama adalah kurangnya objektivitas dan komprehensivitas dalam penilaian, di mana kriteria yang digunakan cenderung sempit, hanya terpaku pada satu aspek dan rentan terhadap subjektivitas penilai [1]. Situasi ini berpotensi menyebabkan ketidakadilan dalam pemberian reward dan mengurangi dampak positif yang seharusnya dapat dihasilkan, sehingga memerlukan Solusi yang lebih terstruktur dan berbasis data.

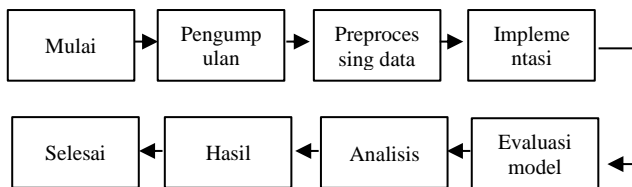
Dalam menghadapi tantangan tersebut, perkembangan teknologi informasi dan ilmu data telah membuka jalan pendekatan yang lebih inovatif. Untuk mengeksplorasi informasi dan pola-pola tersembunyi dalam suatu data, yang dikenal dengan istilah data mining. Salah satu teknik klasifikasi yang populer dalam Data Mining dan telah banyak diterapkan sebagai Solusi dalam berbagai Sistem Pendukung

Keputusan (SPK), termasuk dalam bidang Pendidikan, adalah algoritma klasifikasi K-Nearest Neighbor (KNN). Kajian penelitian sebelumnya telah menunjukkan efektivitas KNN dalam mendukung Keputusan berbasis multikriteria, misalnya studi yang dilakukan oleh Olivia [2], berhasil menerapkan KNN untuk Ketepatan Waktu Lulus Mahasiswa, dalam penelitiannya menjelaskan bahwa penerapan algoritma K-Nearest Neighbor (KNN) untuk memprediksi ketepatan waktu kelulusan mahasiswa berdasarkan data akademik seperti IPK, IPS, SKS, dan lama studi. Demikian pula, penelitian oleh Ma'rif [3] menunjukkan bagaimana KNN dapat digunakan untuk menentukan kelulusan mata pelajaran, penulis menjelaskan bahwa penggunaan algoritma K-Nearest Neighbor (KNN) sebagai solusi untuk memperkirakan dan mengklasifikasikan kelulusan mata pelajaran siswa di tingkat SMA. Prinsip dasar KNN adalah membandingkan jarak antara data uji dengan sejumlah data latih guna menemukan kesamaan terdekat [4][5] di mana data latih yang paling menyerupai data uji berdasarkan kedekatan jarak akan menentukan klasifikasi atau rekomendasi [6]. Dalam penelitian ini, metode KNN digunakan untuk mengolah data calon penerima reward tahun pelajaran ini dengan membandingkan data tersebut dengan data penerima reward dari tahun pelajaran sebelumnya.

II. METODOLOGI PENELITIAN

Penelitian ini dilakukan pada Sekolah Menengah Atas berbasis agama yang memiliki program pemberian reward

pada tiap kelulusan. Pada penelitian ini, langkah dan alur dapat digambarkan sebagai berikut:



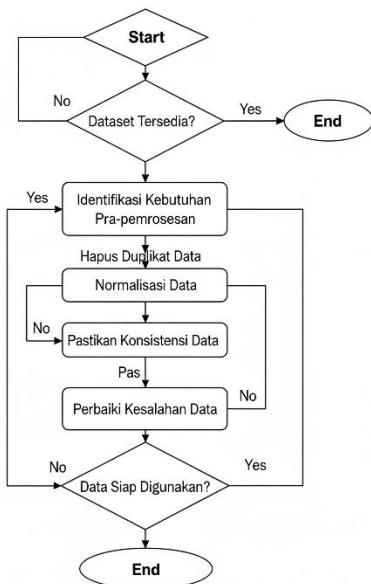
Gambar 1. Alur Tahapan Penelitian

A. Tahapan Pengumpulan Data

Pada penelitian ini, dikumpulkan informasi data siswa lulusan tahun 2024 dan 2025. Dari data tersebut akan dibagi, dengan data lulusan tahun 2024 sebagai data training dan lulusan tahun 2025 sebagai data testing. Dengan melakukan normalisasi data maka didapatkan jumlah data training yang digunakan sebanyak 130 data, sedangkan data testing terdiri dari 76 data. Dalam penelitian ini atribut yang digunakan adalah rata-rata nilai rapor, nilai akhlak, penghasilan orang tua, jumlah saudara dan skor dalam kegiatan non akademik. Kategori yang ditetapkan adalah Layak mendapat reward dan Tidak. Dari data tersebut akan dianalisis menggunakan teknik KNN untuk memberikan rekomendasi kepada sekolah dalam optimasi pemberian reward kepada para siswanya.

B. Tahap Preprocessing Data

Data mentah yang telah dikumpulkan seringkali tidak siap untuk langsung diproses oleh sebuah algoritma. Pada tahap ini bertujuan untuk melakukan pembersihan data dan menyiapkan data [7], prosesnya digambarkan dalam flowchart sebagai berikut:



Gambar 2. Proses Pembersihan dan Pra-pemrosesan

Keterangan gambar:

- 1) *Menghapus duplikat data*: Mengidentifikasi dan menghapus baris data yang sama persis/duplikat dalam dataset.
- 2) *Normalisasi data*: Dengan mengubah skala atribut numerik agar berada dalam rentang yang seragam (misalnya 0-1).

3) *Konsistensi data*: Memastikan format, tipe dan nilai data pada seluruh dataset menjadi seragam.

4) *Memperbaiki kesalahan data*: Penanganan pada nilai yang hilang dan mengoreksi entri yang mungkin salah ketik (typos).

C. Implementasi K-Nearest Neighbor (KNN)

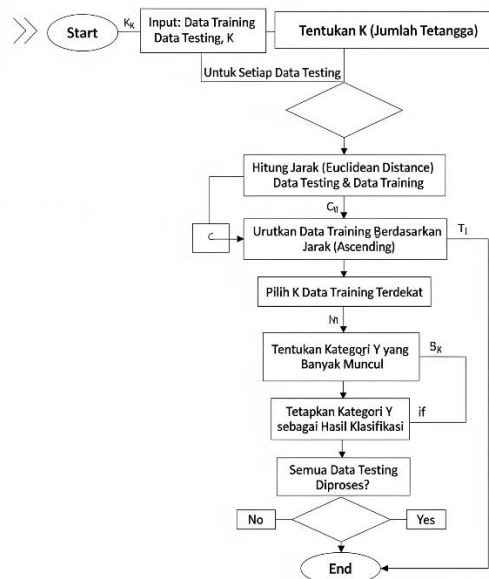
KNN merupakan algoritma yang tergolong dalam metode supervised learning, karena setiap data uji diklasifikasikan berdasarkan kategori terbanyak dari K tetangga terdekatnya. Salah satu keunggulan dari algoritma ini adalah ketahanannya terhadap data latih yang mengandung noise, serta kemampuannya dalam menangani jumlah data yang besar. Di sisi lain, terdapat beberapa kekurangan, seperti keharusan menentukan nilai K secara tepat sebagai jumlah tetangga terdekat. Selain itu, pemilihan jenis jarak dan atribut yang digunakan dapat mempengaruhi akurasi klasifikasi, karena prosesnya memerlukan uji coba untuk memperoleh kombinasi terbaik. Proses ini juga menimbulkan beban komputasi yang tinggi karena jarak antara data uji dan seluruh data latih harus dihitung satu per satu [8]. Untuk keperluan tersebut, biasanya digunakan rumus Euclidean distance seperti yang tercantum pada Persamaan 1[9].

$$d_i = \sqrt{\sum_{i=1}^p (x_1 - x_2)^2} \quad (1)$$

Keterangan :

- d = jarak
- i = variabel data
- p = dimensi data
- x_1 = sampel data
- x_2 = data uji

Alur pemrosesan dengan algoritma K-Nearest Neighbor dalam penelitian ini diuraikan sebagai berikut:



Gambar 3. Proses Pemrosesan Algoritma KNN

Keterangan gambar:

- 1) *Menentukan*: Jumlah tetangga terdekat dengan simbol nilai parameter K.
- 2) *Menghitung*: Jarak antara data training dan data testing dengan menggunakan matriks jarak Euclidean distance dengan rumus persamaan 1.
- 3) *Mengurutkan*: Data berdasarkan hasil nilai Euclidean distance secara ascending atau urut dari terendah ke nilai tinggi.
- 4) *Hasil* : Kategori Y yang muncul paling banyak akan menjadi hasil dari klasifikasi.

D. Tahap Pengujian

Pada tahap ini menjelaskan bagaimana evaluasi dan verifikasi terhadap model. Confusion matrix merupakan sebuah tabel yang berisi hasil uji model klasifikasi. Yang ditampilkan pada Tabel 1 [10].

Tabel 1. Confusion Matrix

	Aktual Positif(1)	Aktual Negatif(0)
Prediksi Positif(1)	TP (True Pofitive)	FP (False Positive)
Prediksi Negatif(0)	FN (False Negative)	TN (True Negative)

Beberapa matrix evaluasi yang sering digunakan untuk menilai kemampuan suatu algoritma klasifikasi ditampilkan pada Tabel 2 [11].

Tabel 2. Matrix Evaluasi

Indikator Uji	Rumus
Akurasi	$\frac{TP + TN}{(TP + TN + FP + FN)}$
Presisi	$\frac{TP}{(TP + FP)}$
Recall	$\frac{TP}{(TP + FN)}$
F1-Score	$2 \times \frac{(Presisi \times Recall)}{(Presisi + Recall)}$
Spesifisitas	$\frac{TN}{(TN + FP)}$

III. HASIL DAN PEMBAHASAN

Data yang digunakan dalam penelitian ini adalah data lulusan siswa tahun 2024 dan 2025 yang akan diolah menggunakan algoritma KNN.

A. Pengumpulan dan Preprocessing Data Awal

Informasi yang disajikan dalam Tabel 3 adalah data awal penerima reward yang belum diolah menggunakan algoritma K-Nearest Neighbor (KNN). Data lulusan tahun 2024 berfungsi sebagai data pelatihan (training data), sementara Tabel 4 memuat data calon penerima reward yang akan diuji (testing data). Kolom nilai rapor pada tabel ini, diisi dengan nilai rata-rata dari semester 1 hingga semester 5. Sedangkan kolom jumlah saudara didapat dari database sekolah. Kemudian penilaian akhlak merujuk pada pengamatan sikap dan kepribadian selama masa studi, yang dikategorikan ke dalam lima tingkat: Sangat Kurang, Kurang, Cukup, Baik, dan Sangat Baik yang ditunjukkan dalam Tabel 5.

Tabel 6 merupakan tabel nilai penghasilan orang tua berdasar pada pendapatan orang tua setiap bulannya. Penilaian kolom ini dibuat 4 parameter yaitu penghasilan diatas 7 juta/bulan, penghasilan 5–7 juta/bulan, penghasilan 2,5–5 juta/bulan dan penghasilan kurang dari 2,5 juta/bulan.

Pengisian kolom skor kegiatan non akademik didasarkan pada nilai keikutsertaan siswa dalam kegiatan-kegiatan non akademik, misalnya aktif dalam organisasi OSIS dan partisipasi dalam lomba-lomba. Parameter skor kegiatan ada 5 yaitu Tidak pernah ikut, Jarang ikut, Kadang-kadang aktif, Aktif (mengikuti kegiatan rutin) dan Aktif (memegang peran penting). Pada penilaian ini ditunjukkan oleh Tabel 7.

Tabel 3. Data Training

No	Nomor Induk Siswa	Rata2 rapor	Nilai Akhlaq	Penghasilan Orangtua	Jumlah saudara	Skor Kegiatan	Mendapat Reward
1	MA-951	85	4	3	1	4	Tidak
2	MA-952	78,33	4	1	6	3	Tidak
3	MA-953	84	4	1	7	4	Tidak
4	MA-954	84,59	4	3	3	4	Tidak
5	MA-955	86,44	4	1	7	4	Tidak
6	MA-956	87,96	4	1	4	3	Tidak
7	MA-957	81,7	4	1	4	3	Tidak
8	MA-958	87	4	4	4	5	Tidak
9	MA-959	87	4	3	3	5	Tidak
10	MA-961	81,85	4	1	3	4	Tidak
11	MA-962	85	4	2	3	4	Tidak
12	MA-963	85,63	4	2	2	4	Tidak
13	MA-964	84	4	1	4	4	Tidak
14	MA-965	86	4	1	4	5	Tidak
15	MA-966	84,41	4	1	4	3	Tidak
16	MA-968	81,93	4	3	3	3	Tidak
17	MA-969	88	4	3	2	5	Tidak
18	MA-970	87	4	1	3	3	Tidak
19	MA-971	89	4	3	4	4	Tidak
20	MA-972	86	4	3	8	3	Tidak
...
130	MA-1073	90,89	4	1	2	3	Reward

Tabel 4. Data Testing

No	NIS	Rata2 rapor	Akhlaq	skor penghasilan	Jumlah saudara	skor kegiatan
1	MA-1134	87,25	4	3	3	5
2	MA-1135	84,33	4	2	3	4
3	MA-1136	84	3	1	2	3
4	MA-1141	84	4	3	2	5
5	MA-1142	85	4	2	5	5
6	MA-1143	86	4	4	3	4
7	MA-1144	89	4	2	3	5
8	MA-1145	87	4	2	3	5
9	MA-1147	84	4	2	2	5
10	MA-1148	77,3	4	1	3	4
11	MA-1149	85	3	2	3	4
12	MA-1150	85,79	4	2	1	4
13	MA-1151	86	4	2	3	3
14	MA-1153	86	4	2	2	5
15	MA-1155	87	5	2	3	5
16	MA-1156	84,49	4	3	2	4
17	MA-1158	84	4	3	5	5
18	MA-1159	85	5	2	3	3
19	MA-1160	84	4	2	2	3

20	MA-1161	89	5	3	5	3
...
76	MA-1239	88,5	4	1	3	3

Tabel 5. Nilai Akhlak

Parameter	Nilai
Sangat kurang	1
Kurang	2
Cukup	3
Baik	4
Sangat baik	5

Tabel 6. Penghasilan Orang Tua

Parameter	Nilai
> 7.000.000	1
5.000.001 - 7.000.000	2
2.500.001 - 5.000.000	3
< 2.500.000	4

Tabel 7. Partisipasi Kegiatan

Parameter	Nilai
Tidak pernah ikut	1
Jarang ikut	2
Kadang-kadang aktif	3
Aktif (mengikuti dan berkontribusi secara rutin)	4
Aktif (rutin mengikuti, memegang peran penting)	5

B. Implementasi Algoritma KNN

1) *Perhitungan KNN:* Berikut ini akan dijelaskan langkah proses perhitungan dengan metode KNN berdasar data training dan dan testing diatas, dengan menggunakan rumus persamaan 1. Data testing yang digunakan sebagai contoh adalah data dengan Nomor Induk Siswa MA-1134 adalah sebagai berikut:

$$d_i = \sqrt{(85 - 87,25)^2 + (4 - 4)^2 + (3 - 3)^2 + (1 - 3)^2 + (4 - 5)^2}$$

$$d_i = 3,17$$

Tabel 8. Hasil perhitungan Euclidean Distance

No	Nomor Induk Siswa	Rata2 rapor	Nilai Akhlaq	Penghasilan Orangtua	Jumlah saudara	Skor Kegiatan	Nilai Euclidean distance
1	MA-951	85	4	3	1	4	3,17
2	MA-952	78,33	4	1	6	3	9,83
3	MA-953	84	4	1	7	4	5,62
4	MA-954	84,59	4	3	3	4	2,84
5	MA-955	86,44	4	1	7	4	4,65
6	MA-956	87,96	4	1	4	3	3,08
7	MA-957	81,7	4	1	4	3	6,31
8	MA-958	87	4	4	4	5	1,44
9	MA-959	87	4	3	3	5	0,25
10	MA-961	81,85	4	1	3	4	5,84
11	MA-962	85	4	2	3	4	2,66
12	MA-963	85,63	4	2	2	4	2,37
13	MA-964	84	4	1	4	4	4,07
14	MA-965	86	4	1	4	5	2,56
15	MA-966	84,41	4	1	4	3	4,13
16	MA-968	81,93	4	3	3	3	5,68
17	MA-969	88	4	3	2	5	1,25
18	MA-970	87	4	1	3	3	2,84
19	MA-971	89	4	3	4	4	2,25
20	MA-972	86	4	3	8	3	5,53
...
130	MA-1073	90,89	4	1	2	3	4,72

Tabel 9. Sort Ascending Nilai Euclidean Distance

No	Nomor Induk Siswa	Rata2 rapor	Nilai Akhlaq	Penghasilan Orangtua	Jumlah saudara	Skor Kegiatan	Nilai Euclidean distance	reward
1	MA-959	87	4	3	3	5	0,25	Tidak
2	MA-969	88	4	3	2	5	1,25	Tidak
3	MA-958	87	4	4	4	5	1,44	Tidak
4	MA1000	87	4	2	2	4	1,75	Tidak
5	MA1057	88	4	2	4	4	1,89	Tidak
6	MA1082	88,56	4	3	4	4	1,93	Tidak
7	MA1006	85,37	4	3	4	5	2,13	Tidak
8	MA-980	88	4	3	3	3	2,14	Tidak
9	MA1022	86	4	4	2	4	2,14	Tidak
10	MA-971	89	4	3	4	4	2,25	Tidak
11	MA1034	87	4	1	3	4	2,25	Tidak
12	MA1001	86,93	4	1	3	4	2,26	Tidak
13	MA1060	87,7	4	3	2	3	2,28	Tidak
14	MA1009	86,51	4	1	4	5	2,36	Reward
15	MA-987	88	4	1	4	5	2,36	Tidak
16	MA-963	85,63	4	2	2	4	2,37	Tidak
17	MA-965	86	4	1	4	5	2,56	Tidak
18	MA-975	86	4	3	2	3	2,56	Tidak
19	MA-981	88	4	1	4	4	2,56	Tidak
20	MA-990	86	4	2	3	3	2,56	Tidak
...
130	MA1021	90,89	4	1	2	3	4,72	Tidak

Berdasarkan hasil perhitungan Euclidean distance yang ditunjukkan dalam Tabel 8, dan setelah data diurutkan menurut kedekatan jaraknya sesuai nilai K pada Tabel 9, proses klasifikasi dilakukan untuk menentukan apakah seorang siswa berhak menerima reward atau tidak. Dalam kasus data testing nomor 1 dengan K = 5 dan NIS MA-1134, yang memiliki rata-rata nilai rapor sebesar 7,25, nilai akhlak 4, penghasilan orang tua 3, jumlah saudara 3, serta skor kegiatan 5, diperoleh hasil bahwa seluruh lima tetangga terdekat berasal dari data latih yang tidak menerima reward. Dengan demikian, siswa MA-1134 dikategorikan sebagai tidak layak menerima reward karena mayoritas data pembandingan menunjukkan hasil serupa.

Selanjutnya adalah perhitungan dari keseluruhan data testing yang ada. Dari data testing sebanyak 76 data dengan K=5, ternyata didapatkan hasil “Tidak layak mendapatkan Reward” untuk keseluruhan data. Karena dalam penelitian ini bertujuan untuk memilih siswa yang dikategorikan “Layak Mendapat Reward” maka akan ditentukan nilai K yang lebih kecil, dengan tujuan lebih sensitif dengan terhadap banyaknya data training dengan kategori “Layak Mendapat Reward” yang berjumlah sedikit. Dari hasil perhitungan didapat hasil seperti tabel sebagai berikut:

Tabel 10. Nilai Euclidean dengan berbagai nilai K

No	Kategori	K = 5	K = 3	K = 1
1.	Layak Mendapat Reward	0	4	8
2.	Tidak layak mendapat Reward	76	72	68

2) *Implementasi Sistem:* Implementasi sistem adalah proses penerapan studi kasus ke dalam bentuk program pada perangkat lunak. Dalam hal ini program menggunakan bahasa python. Ditunjukkan pada gambar 2.

```
import pandas as pd
from sklearn.preprocessing import MinMaxScaler, LabelEncoder

# 1. Load Dataset
# Ganti 'data.csv' dengan file dataset Anda
data = pd.read_csv("data.csv")

...

# 7. Menyimpan Dataset yang Sudah Diproses
# Simpan data untuk digunakan dalam algoritma KNN
data.to_csv("processed_data.csv", index=False)
print("\nDataset yang sudah diproses disimpan ke 'processed_data.csv'.")
```

```
from sklearn.model_selection import train_test_split
from sklearn.neighbors import KNeighborsClassifier
from sklearn.metrics import accuracy_score, precision_score, recall_score,
f1_score, classification_report, confusion_matrix
import matplotlib.pyplot as plt
import seaborn as sns

# 1. Load Preprocessed Data
# Ganti dengan file dataset yang sudah diproses
data = pd.read_csv("processed_data.csv")

...
```

Gambar 2. Kode Python

3) *Pengujian Metode KNN*: Pada penelitian ini telah dilakukan evaluasi metode KNN dengan menghitung nilai akurasi metode KNN dengan metode confusion matrix. Hasil dari pengujian ini dapat dilihat pada tabel 11 dan tabel 12.

Tabel 11. Matrix Evaluasi

Indikator Uji	Hasil	Persen
Akurasi	0,907894737	91%
Presisi	0,666666667	67%
Recall	0,444444444	44%
F1-Score	0,533333333	53%
Spesifisitas	0,970149254	97%

Tabel 12. Confusion Matrix

n=76 (jumlah data testing)	Aktual Positif(1)	Aktual Negatif(0)	Aktual Positif(1)	Aktual Negatif(0)
Prediksi Positif(1)	TP	FP	4	2
Prediksi Negatif(0)	FN	TN	5	65

Tabel 12 merupakan hasil dari pengujian yang didapatkan menggunakan confusion matrix yang diinterpretasikan dalam kategori sebagai berikut:

- TP** = True Positive sebanyak 4 data yang diprediksi mendapat reward dan benar
- TN** = True Negative sebanyak 65 data yang diprediksi tidak mendapat reward dan benar
- FP** = False Positive sebanyak 2 data yang diprediksi mendapat reward dan salah
- FN** = False Negative sebanyak 5 data yang diprediksi tidak mendapat reward dan salah

IV. KESIMPULAN

Dalam penelitian ini dikatakan berhasil dalam mengimplementasikan algoritma K-Nearest Neighbor (KNN) sebagai metode untuk mengoptimalkan sistem pemberian reward siswa di tingkat SMA. Dengan mempertimbangkan multikriteria seperti nilai rata-rata rapor, tingkat penghasilan orang tua, jumlah saudara, nilai akhlak dan partisipasi dalam

kegiatan non akademik. Dari serangkaian pengujian yang dilakukan, nilai K=5 terindikasi sebagai parameter optimal yang memberikan akurasi keseluruhan tertinggi pada data validasi, yaitu mencapai 91%.

Namun, setelah melalui proses pengujian dengan confusion matrix, analisis menunjukkan bahwa model condong terhadap kelas mayoritas ("Tidak Layak Mendapat Reward") akibat ketidakseimbangan data, sehingga kurang optimal dalam mengidentifikasi siswa yang "Layak Mendapat Reward" pada K=5. Meskipun demikian, eksplorasi K yang lebih kecil (seperti K=1 dan K=3) menunjukkan potensi dalam mendeteksi kelas minoritas, meskipun dengan penurunan akurasi keseluruhan. Nilai F1-score yang tidak begitu tinggi juga dipengaruhi oleh sifat model yang dirancang untuk bersifat selektif, yaitu hanya merekomendasikan siswa yang benar-benar memenuhi kriteria tertinggi untuk menerima reward. Meskipun recall rendah menunjukkan bahwa tidak semua siswa yang layak berhasil terdeteksi, presisi yang relatif lebih baik mengindikasikan bahwa sebagian besar yang direkomendasikan memang tepat sasaran. Pendekatan ini disesuaikan dengan konteks pemberian reward yang kuotanya terbatas, sehingga sistem lebih mengutamakan ketepatan dibandingkan kuantitas.

V. SARAN

Implikasi dari temuan penelitian ini adalah meskipun KNN dapat menyajikan kerangka kerja yang terstruktur dan objektif untuk evaluasi multikriteria, tetapi dengan dataset yang tidak seimbang dapat membatasi efektivitasnya dalam mengidentifikasi kelas minoritas. Oleh karena itu, untuk penelitian di masa yang akan datang, disarankan untuk mengeksplorasi teknik-teknik penanganan ketidakseimbangan dataset, seperti oversampling atau undersampling. Serta menggunakan algoritma klasifikasi lain yang mungkin lebih kokoh terhadap ketidakseimbangan data.

REFERENSI

- [1] S. P. Abduloh, M. P. Suntoko, S. P. Tedi Purbangkara, M. P. Ade Abikusna, and others, *Peningkatan dan Pengembangan Prestasi Belajar Peserta Didik*. Ponorogo: Uwais Inspirasi Indonesia, 2022.
- [2] D. Olivia, "Penerapan Algoritma K-Nearest Neighbor (KNN) Untuk Ketepatan Waktu Lulus Mahasiswa," Makassar, 2024. (Skripsi, Universitas Muhammadiyah Makassar, 2024). Tanggal akses 18 April 2025, Diakses dari https://digilibadmin.unismuh.ac.id/upload/39413-Full_Text.pdf
- [3] M. Ma'ruf, "Penerapan Algoritma K-Nearest Neighbor Untuk Menentukan Kelulusan Mata Pelajaran Di Sma N 1 Semarang". (Skripsi, Universitas Semarang, 2024). Tanggal Akses 18 April 2025, Diakses dari <https://eskripsi.usm.ac.id/files/skripsi/G11A/2020/G.131.20.0002>
- [4] F. M. D. Maharani, A. L. Hananto, S. S. Hilabi, F. N. Apriani, A. Hananto, and B. Huda, "Perbandingan Metode Klasifikasi Sentimen Analisis Penggunaan E-

- Wallet Menggunakan Algoritma Naïve Bayes dan K-Nearest Neighbor,” *Metik Jurnal*, vol. 6, no. 2, pp. 97–103, 2022.
- [5] S. D. Prasetyo, S. S. Hilabi, and F. Nurapriani, “Analisis Sentimen Relokasi Ibukota Nusantara Menggunakan Algoritma Naïve Bayes dan KNN,” *Jurnal KomtekInfo*, pp. 1–7, 2023.
- [6] H. Andriana, “Penerapan Metode K-Nearest Neighbor pada Sentimen Analisis Pengguna Twitter Terhadap KTT G20 di Indonesia,” *JIKO (Jurnal Inform. dan Komputer)*, 2023.
- [7] A. Tawakuli, B. Havers, V. Gulisano, D. Kaiser, and T. Engel, “Survey: Time-series data preprocessing: A survey and an empirical analysis,” *Journal of Engineering Research*, 2024.
- [8] S. R. Cholil, T. Handayani, R. Prathivi, and T. Ardianita, “Implementasi algoritma klasifikasi k-nearest neighbor (knn) untuk klasifikasi seleksi penerima beasiswa,” *IJCIT (Indonesian Journal on Computer and Information Technology)*, vol. 6, no. 2, pp. 118–127, 2021.
- [9] W. Wang *et al.*, “Euclidean distance-based Raman spectroscopy (EDRS) for the prognosis analysis of gastric cancer: A solution to tumor heterogeneity,” *Spectrochim Acta A Mol Biomol Spectrosc*, vol. 288, p. 122163, 2023.
- [10] M. A. Aprihartha, J. Prasetya, and S. I. Fallo, “Implementasi CART-Real Adaboost dalam Memprediksi Minat Pelanggan Membeli Sepatu,” *Jurnal EurekaMatika*, vol. 12, no. 1, pp. 35–46, 2024.
- [11] M. A. Aprihartha, T. N. Alam, and M. Husniyadi, “Perbandingan Metrik Euclidean dan Metrik Manhattan untuk K-Nearest Neighbors dalam Klasifikasi Kismis,” *Jurnal Ilmu Komputer dan Informatika*, vol. 4, no. 1, pp. 21–30, 2024.