

Klasifikasi Penerimaan Peserta Didik Baru Berdasarkan Sistem Zonasi Menggunakan Algoritma K-Nearest Neighbors

Muhammad Syah Fiqri*, **Andi Taufik**

Fakultas Teknik dan Informatika, Program Studi Informatika, Universitas Bina Sarana Informatika, Jakarta, Indonesia

Email: ^{1,*}syachfiqri@gmail.com, ¹andi.iuf@bsi.ac.id

Email Penulis Korespondensi: syachfiqri@gmail.com

Abstrak—Penerapan sistem zonasi pada Penerimaan Peserta Didik Baru (PPDB) sering menimbulkan permasalahan dalam menentukan kelayakan siswa berdasarkan domisili, sehingga diperlukan pendekatan berbasis data untuk mendukung proses seleksi. Penelitian ini bertujuan mengklasifikasikan siswa baru SMPN 16 Bogor tahun ajaran 2025 dengan menggunakan algoritma K-Nearest Neighbors (KNN). Dataset yang digunakan terdiri dari 1.153 entri siswa dengan atribut longitude, latitude, jarak rumah ke sekolah, serta label zonasi dan nonzonasi. Tahap preprocessing dilakukan melalui data cleaning, label encoding, dan standardization sebelum data dibagi menjadi 75% data latih dan 25% data uji. Nilai K optimal ditentukan pada K=18 dengan error rate minimum 0,1591. Hasil pengujian menunjukkan akurasi 97% pada data latih dan 84% pada data uji, yang mengindikasikan model bekerja dengan cukup baik meskipun terdapat indikasi overfitting. Penelitian ini berkontribusi dengan membuktikan bahwa atribut spasial dapat diintegrasikan secara efektif ke dalam sistem klasifikasi PPDB, serta memberikan dasar bagi pengembangan sistem pendukung keputusan yang lebih objektif dan adaptif di masa depan.

Kata Kunci: Klasifikasi; K-Nearest Neighbors; Sistem Zonasi; PPDB

Abstract—The implementation of the zoning system in student admission (PPDB) often raises challenges in determining eligibility based on domicile, requiring a data-driven approach to support the selection process. This study aims to classify new students of SMPN 16 Bogor for the 2025 academic year using the K-Nearest Neighbors (KNN) algorithm. The dataset consists of 1,153 student records with attributes including longitude, latitude, distance from home to school, and zoning labels. Preprocessing involved data cleaning, label encoding, and feature standardization before splitting the data into 75% training and 25% testing sets. The optimal parameter was found at K=18 with a minimum error rate of 0.1591. Experimental results showed an accuracy of 97% for training data and 84% for testing data, indicating that the model performs reasonably well despite signs of overfitting. This research contributes by demonstrating that spatial attributes can be effectively integrated into zoning-based classification and provides a foundation for developing more objective and adaptive decision support systems in the context of student admissions.

Keywords: Classification; K-Nearest Neighbors; Zoning System; Student Admission

1. PENDAHULUAN

Sekolah merupakan lembaga pendidikan yang berperan penting dalam mempersiapkan generasi muda menghadapi masa depan. Salah satu agenda rutin yang dilakukan setiap tahun ajaran baru di seluruh jenjang pendidikan formal adalah penerimaan peserta didik baru (PPDB). Kebijakan mengenai PPDB di Indonesia mengalami perubahan signifikan sejak diberlakukannya sistem zonasi oleh Kementerian Pendidikan dan Kebudayaan melalui Permendikbud Nomor 51 Tahun 2018. Regulasi tersebut kemudian dievaluasi dalam berbagai penelitian yang menyoroti implementasi PPDB berbasis zonasi, di mana prinsip utama penerimaan peserta didik diutamakan berdasarkan kedekatan jarak domisili dengan sekolah [1]. Tujuannya adalah untuk mendorong pemerataan kualitas pendidikan dengan mengurangi kesenjangan antara sekolah unggulan dan sekolah non-unggulan.

Namun, implementasi sistem zonasi tidak selalu berjalan mulus. Beberapa sekolah menghadapi kendala berupa ketidakseimbangan antara kapasitas penerimaan dengan jumlah pendaftar, perbedaan jarak antar sekolah yang cukup jauh, hingga perbedaan kepadatan penduduk di tiap wilayah. Akibatnya, muncul keresahan baik di kalangan orang tua maupun sekolah terkait dengan efektivitas sistem zonasi dalam menjamin keadilan dan pemerataan pendidikan. Dalam konteks inilah, diperlukan suatu pendekatan berbasis teknologi yang mampu membantu sekolah melakukan klasifikasi calon peserta didik sesuai ketentuan zonasi secara objektif dan terukur.

Salah satu pendekatan yang dapat digunakan adalah metode data mining, khususnya algoritma K-Nearest Neighbors (KNN). KNN merupakan algoritma klasifikasi sederhana namun efektif, yang bekerja dengan prinsip mengklasifikasikan suatu data baru berdasarkan kedekatan dengan sejumlah data latih [2]. Nilai K mewakili jumlah tetangga terdekat yang digunakan dalam proses klasifikasi, sementara ukuran kedekatan umumnya dihitung menggunakan Euclidean distance.[3] Metode ini termasuk non-parametrik dan berbasis memori, sehingga cocok digunakan dalam kasus PPDB yang melibatkan data spasial seperti longitude dan latitude.

KNN telah banyak digunakan dalam penelitian, seperti yang dilakukan oleh Muhammin et al [2], mengkaji klasifikasi prestasi akademik berdasarkan nilai rapor dan kedisiplinan siswa. Nawar et al [4], mengembangkan metode Weighted Product (WP) untuk penilaian soft skill guru, sementara Imani[5] menerapkan KNN untuk rekomendasi keminatan studi di perguruan tinggi. Kartika & Santoso[6] memanfaatkan kombinasi KNN dan WP untuk seleksi siswa berprestasi, sedangkan Widaningsih[7] serta Samudra et al [8], mengaplikasikan KNN pada prediksi kelulusan mahasiswa. Di sisi lain, penelitian Wati[9], Putri et al [10], dan Nasution[11] juga menekankan efektivitas KNN dalam memprediksi hasil akademik mahasiswa. Di bidang kesehatan, Henderi et al [12], membandingkan normalisasi Min-Max dan Z-Score pada KNN untuk klasifikasi jenis kanker payudara.

Dalam konteks PPDB, penelitian terbaru menunjukkan bahwa KNN dapat membantu pengelompokan calon peserta didik berdasarkan zonasi secara lebih objektif. Misalnya, Muhammin et al.[2] menyoroti efektivitas KNN dalam klasifikasi berbasis data pendidikan, sementara studi Darya[1] menunjukkan bagaimana kebijakan PPDB zonasi masih memerlukan dukungan teknologi agar implementasinya lebih adil dan efisien. Meski demikian, kajian yang secara khusus mengintegrasikan atribut spasial (longitude dan latitude) pada kasus PPDB tingkat sekolah menengah pertama masih jarang dilakukan.

Perbedaan inilah yang menjadi GAP penelitian ini. Jika penelitian sebelumnya cenderung fokus pada aspek prestasi siswa atau prediksi kelulusan, penelitian ini menghadirkan kontribusi baru dengan menjadikan atribut spasial sebagai parameter utama klasifikasi. Misalnya, berbeda dengan Kartika & Santoso[6] yang menitikberatkan analisis pada performa akademik, penelitian ini menggunakan lokasi domisili peserta didik sebagai variabel utama dalam penentuan zonasi. Demikian pula, dibandingkan studi evaluasi PPDB yang lebih menekankan aspek kebijakan,[2] penelitian ini menambahkan koordinat longitude dan latitude agar klasifikasi lebih akurat. Dengan pendekatan ini, diharapkan sistem zonasi dapat dipetakan secara lebih objektif, sekaligus memberikan nilai tambah dalam implementasi kebijakan PPDB.

Selain mengatasi kendala praktis dalam pelaksanaan zonasi, penelitian ini juga berkontribusi secara teoretis. Pertama, penelitian ini memperluas penerapan KNN pada ranah pendidikan dengan integrasi data spasial. Kedua, penelitian ini menguji efektivitas zonasi di tingkat sekolah menengah pertama dengan dataset nyata dari SMPN 16 Bogor, sehingga hasilnya relevan dengan praktik lapangan. Ketiga, penelitian ini mengukur akurasi model klasifikasi berbasis KNN melalui evaluasi confusion matrix dan classification report menggunakan metrik evaluasi terbaru.[13] Dengan kontribusi tersebut, penelitian ini diharapkan tidak hanya memberikan solusi praktis bagi pihak sekolah dalam menyaring calon peserta didik baru sesuai sistem zonasi, tetapi juga memperkaya literatur akademik mengenai penerapan algoritma klasifikasi dalam bidang pendidikan.

2. METODOLOGI PENELITIAN

Penelitian ini dilaksanakan di SMPN 16 Bogor dengan memanfaatkan data penerimaan peserta didik baru tahun ajaran 2025 yang berjumlah 1.153 entri. Setiap entri mencakup informasi identitas, jenis kelamin, koordinat longitude dan latitude tempat tinggal siswa, serta jarak rumah ke sekolah, dengan label klasifikasi berupa zonasi dan nonzonasi. Tahap pertama dalam metodologi adalah pengumpulan data yang bersumber dari dokumentasi resmi sekolah, sehingga menjamin validitas dataset yang digunakan. Data tersebut kemudian diproses melalui tahapan preprocessing agar siap diolah oleh algoritma K-Nearest Neighbors (KNN).

Proses preprocessing dilakukan secara sistematis. Pada tahap awal dilakukan pengecekan nilai hilang untuk memastikan kelengkapan dataset. Hasil pemeriksaan menunjukkan tidak ada missing value, sehingga semua data dapat dipertahankan. Selanjutnya, data kategorikal berupa label zonasi dan nonzonasi diubah menjadi bentuk numerik menggunakan teknik label encoding, di mana zonasi diberi kode 1 dan nonzonasi diberi kode 0. Transformasi ini penting agar data dapat diolah secara matematis oleh KNN. Setelah itu, dilakukan normalisasi menggunakan StandardScaler agar skala antar atribut menjadi seragam. Normalisasi dibutuhkan karena variabel longitude dan latitude memiliki rentang nilai yang berbeda, sehingga tanpa proses standarisasi perhitungan jarak Euclidean berpotensi bias. Selain itu, atribut jarak rumah ke sekolah tetap dipertahankan dan dihitung berdasarkan koordinat spasial menggunakan rumus Euclidean distance, di mana semakin kecil jarak yang dihasilkan maka semakin besar kemungkinan siswa termasuk ke dalam kategori zonasi [9][11].

Setelah preprocessing, data dibagi menjadi data latih dan data uji dengan perbandingan 75% untuk data latih dan 25% untuk data uji. Proporsi ini dipilih untuk memberikan keseimbangan antara jumlah data yang digunakan untuk membangun model dan jumlah data yang digunakan untuk menguji kinerja model. Algoritma KNN kemudian diterapkan dengan mencari nilai K optimal melalui pengujian berbagai nilai K dari 1 hingga 20. Hasil eksplorasi menunjukkan bahwa nilai K terbaik berada pada K=18 dengan error terkecil sebesar 0,1591, yang menandakan bahwa pada titik tersebut akurasi model paling stabil [11][8].

Tahap terakhir adalah evaluasi model. Evaluasi dilakukan dengan dua pendekatan, yaitu confusion matrix dan classification report. Confusion matrix dipilih karena dapat menampilkan distribusi prediksi model secara rinci melalui parameter true positive, true negative, false positive, dan false negative. Hal ini penting dalam konteks zonasi, sebab kesalahan klasifikasi dapat berdampak pada penerimaan siswa yang seharusnya ditolak atau diterima. Sementara itu, classification report digunakan untuk mengukur kinerja model secara kuantitatif melalui akurasi, presisi, recall, dan F1-score. Akurasi menunjukkan tingkat ketepatan keseluruhan prediksi, presisi menunjukkan seberapa tepat model mengidentifikasi siswa zonasi, recall menggambarkan kemampuan mendeteksi seluruh siswa yang seharusnya masuk zonasi, dan F1-score memberikan keseimbangan antara presisi dan recall.[9][10][8] Dengan menggunakan kedua metrik ini, hasil penelitian tidak hanya menampilkan tingkat keberhasilan model secara numerik, tetapi juga relevan secara praktis terhadap implementasi kebijakan zonasi di sekolah.

3. HASIL DAN PEMBAHASAN

3.1 Analisis Metode *K-Nearest Neighbors* (KNN)

Dataset yang digunakan dalam penelitian ini terdiri atas 1.153 entri siswa baru SMPN 16 Bogor tahun ajaran 2025. Atribut yang digunakan mencakup NIPD, jenis kelamin, koordinat latitude dan longitude, jarak rumah ke sekolah, serta label zonasi. Data tersebut digunakan untuk melatih dan menguji model KNN dengan rasio pembagian 75% data latih dan 25% data uji.. Tabel 1 berikut memberikan gambaran sebagian isi dataset yang digunakan dalam penelitian.

Tabel 1. Dataset

No	NIPD	Jenis kelamin	Latitude	Longitude	Jarak rumah ke sekolah (KM)	Zonasi
1.	242507001	L	-6.5359	106.7748	1	Zonasi
2.	242507040	L	-6.535653	106.7668	1	Zonasi
3.	242507079	L	-6.535759	106.76675	1	Zonasi
4.	222307338	L	-6.566474247	106.8010139	3	Nonzonasi
5.	232407075	L	-6.5487	106.7622	1	Zonasi
6.	232407037	L	-6.582	106.794	1	Zonasi
7.	222307189	L	-6.53236603	106.7900276	2	Nonzonasi
8.	222307001	L	-6.534246	106.792837	0	Zonasi
9.	232407111	L	-6.516489612	106.7689777	1	Zonasi
...		
1153	222307379	L	-6.63485417	106.7767239	0	Zonasi

3.2 Data Preprocessing

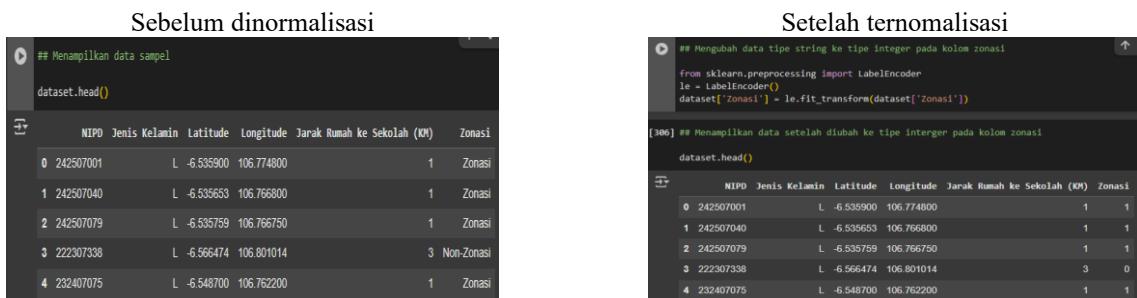
Tahap preprocessing dilakukan untuk memastikan kualitas dataset sebelum masuk ke proses pemodelan. Pada tahap data cleaning, dilakukan pengecekan terhadap missing value untuk memastikan tidak ada data yang hilang. Hasil pengecekan pada Gambar 2 menunjukkan bahwa dataset telah bersih dan dapat digunakan tanpa proses imputasi tambahan.



```
dataset.isnull().sum()
NIPD      0
Jenis Kelamin      0
Latitude      0
Longitude      0
Jarak Rumah ke Sekolah (KM)      0
Zonasi      0
```

Gambar 2. Check missing value

Selanjutnya dilakukan data transformation yang terdiri dari dua langkah utama. Pertama, dilakukan proses label encoding dengan memanfaatkan fungsi LabelEncoder dari library sklearn.preprocessing. Proses ini mengubah variabel kategorikal zonasi menjadi bentuk numerik, di mana label “Zonasi” direpresentasikan dengan angka 1 dan label “Nonzonasi” direpresentasikan dengan angka 0. Perubahan tersebut ditunjukkan pada Gambar 3.



Sebelum dinormalisasi

```
## Menampilkan data sampel
dataset.head()

NIPD Jenis Kelamin Latitude Longitude Jarak Rumah ke Sekolah (KM) Zonasi
0 242507001 L -6.535900 106.774800 1 Zonasi
1 242507040 L -6.535653 106.766800 1 Zonasi
2 242507079 L -6.535759 106.766750 1 Zonasi
3 222307338 L -6.566474 106.801014 3 Non-Zonasi
4 232407075 L -6.548700 106.762200 1 Zonasi
```

Setelah ternormalisasi

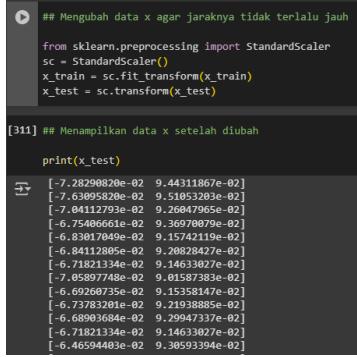
```
## Mengubah data tipe string ke tipe integer pada kolom zonasi
from sklearn.preprocessing import LabelEncoder
le = LabelEncoder()
dataset['Zonasi'] = le.fit_transform(dataset['Zonasi'])

## Menampilkan data setelah diubah ke tipe integer pada kolom zonasi
dataset.head()

NIPD Jenis Kelamin Latitude Longitude Jarak Rumah ke Sekolah (KM) Zonasi
0 242507001 L -6.535900 106.774800 1 1
1 242507040 L -6.535653 106.766800 1 1
2 242507079 L -6.535759 106.766750 1 1
3 222307338 L -6.566474 106.801014 3 0
4 232407075 L -6.548700 106.762200 1 1
```

Gambar 3. Transformasi tipe data *string* ke tipe *integer*

Kedua, dilakukan standarisasi nilai fitur menggunakan fungsi StandardScaler dari library sklearn.preprocessing. Langkah ini penting untuk menghilangkan ketimpangan skala antar atribut, terutama karena fitur longitude, latitude, dan jarak rumah memiliki rentang nilai yang berbeda. Hasil transformasi nilai setelah standarisasi ditampilkan pada Gambar 4.



```

## Mengubah data x agar jaraknya tidak terlalu jauh
from sklearn.preprocessing import StandardScaler
sc = StandardScaler()
x_train = sc.fit_transform(x_train)
x_test = sc.transform(x_test)

## Menampilkan data x setelah diubah
print(x_test)

```

[311] [-7.282008e-02 9.44311867e-02]
[-7.6395920e-02 9.51053203e-02]
[-7.04112793e-02 9.256047965e-02]
[-6.5406651e-02 9.360700765e-02]
[-6.83017049e-02 9.15742119e-02]
[-6.84112895e-02 9.20828427e-02]
[-6.71821134e-02 9.14613027e-02]
[-7.05897748e-02 9.01587383e-02]
[-6.69269735e-02 9.15358147e-02]
[-6.73783201e-02 9.21938885e-02]
[-6.68903684e-02 9.29947337e-02]
[-6.71821334e-02 9.14633827e-02]
[-6.46594403e-02 9.30593394e-02]

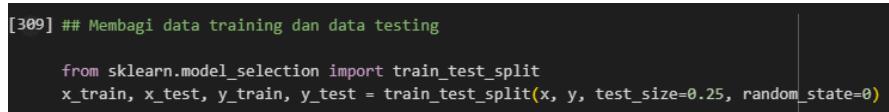
Gambar 4. Transformasi *values* data

Dengan demikian, dataset akhir yang diperoleh sudah bersih dari missing value, label kategorikal telah diubah menjadi numerik, serta seluruh atribut numerik memiliki skala yang seimbang. Hal ini memastikan model KNN dapat bekerja secara optimal, karena algoritma ini sangat bergantung pada perhitungan jarak antar data sehingga rentang fitur yang seimbang menjadi syarat penting dalam proses klasifikasi.

3.3 Data Modeling

Tahap data modeling dilakukan melalui tiga langkah utama: data splitting, penentuan nilai K, serta penerapan algoritma K-Nearest Neighbors (KNN) dengan jarak Euclidean.

Pertama, dataset dibagi menjadi data latih (75%) dan data uji (25%) menggunakan fungsi `train_test_split` dari library `sklearn`. Jumlah data latih adalah 864 entri, sementara data uji terdiri dari 289 entri. Proses pembagian data ini ditunjukkan pada Gambar 5.



```

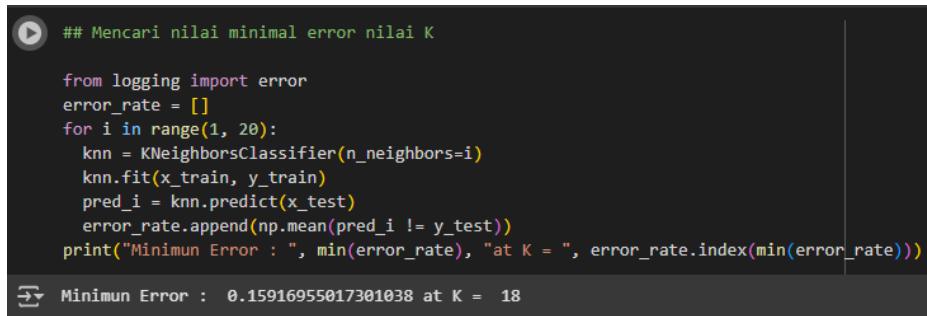
[309] ## Membagi data training dan data testing

from sklearn.model_selection import train_test_split
x_train, x_test, y_train, y_test = train_test_split(x, y, test_size=0.25, random_state=0)

```

Gambar 5. Splitting data *train* & data *test*

Kedua, untuk menentukan nilai K terbaik, dilakukan pengujian terhadap rentang K antara 1 hingga 20 dengan menghitung error rate. Hasil pengujian pada Gambar 6 menunjukkan bahwa nilai K=18 menghasilkan error rate minimum sebesar 0,1591. Nilai ini dipilih karena mampu menyeimbangkan bias dan variansi, sehingga model tidak terlalu sensitif terhadap noise namun tetap mampu menangkap pola distribusi data secara akurat.



```

## Mencari nilai minimal error nilai K

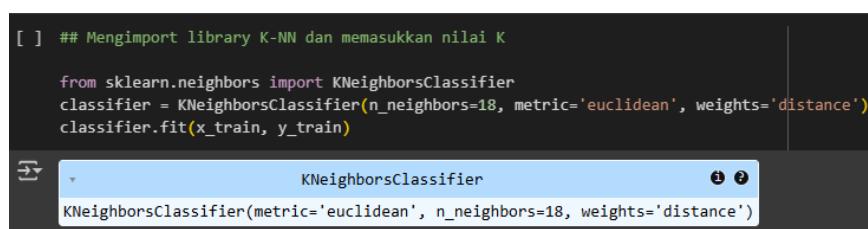
from logging import error
error_rate = []
for i in range(1, 20):
    knn = KNeighborsClassifier(n_neighbors=i)
    knn.fit(x_train, y_train)
    pred_i = knn.predict(x_test)
    error_rate.append(np.mean(pred_i != y_test))
print("Minimun Error : ", min(error_rate), "at K = ", error_rate.index(min(error_rate)))

```

Minimun Error : 0.15916955017301038 at K = 18

Gambar 6. Mencari nilai K paling optimal

Ketiga, pemodelan dilakukan dengan fungsi `KNeighborsClassifier` pada library `sklearn`, menggunakan nilai K terbaik yang telah diperoleh. Perhitungan jarak antar data dilakukan dengan rumus Euclidean (Gambar 7), yang umum digunakan pada KNN karena mampu merepresentasikan kedekatan spasial antar titik dengan baik. Pemilihan metrik ini sejalan dengan karakteristik dataset penelitian yang berbasis koordinat geografis.



```

[ ] ## Mengimport library K-NN dan memasukkan nilai K

from sklearn.neighbors import KNeighborsClassifier
classifier = KNeighborsClassifier(n_neighbors=18, metric='euclidean', weights='distance')
classifier.fit(x_train, y_train)

```

KNeighborsClassifier
KNeighborsClassifier(metric='euclidean', n_neighbors=18, weights='distance')

Gambar 7. Algoritma *K-Nearest Neighbors* (KNN) dengan rumus *euclidean distance*

Dengan demikian, model KNN yang digunakan dalam penelitian ini dibangun melalui pembagian data yang proporsional, pemilihan nilai K optimal, serta penerapan jarak Euclidean untuk menghitung kedekatan antar data. Proses ini menghasilkan model klasifikasi yang siap dievaluasi pada tahap berikutnya.

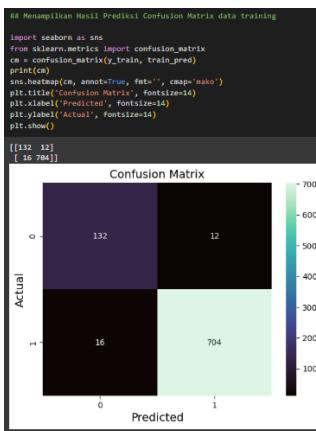
3.4 Evaluasi

Tahap evaluasi dilakukan untuk mengukur performa model KNN dengan memanfaatkan confusion matrix dan classification report. Dua metrik ini dipilih karena mampu memberikan gambaran menyeluruh terkait akurasi, presisi, recall, dan f1-score pada masing-masing kelas.

Hasil evaluasi confusion matrix pada data latih (Tabel 2, Gambar 8) menunjukkan bahwa model mampu mengklasifikasikan data dengan tingkat akurasi 97%. Nilai True Positive (TP) dan True Negative (TN) relatif tinggi, sementara False Positive (FP) dan False Negative (FN) rendah. Kondisi ini menandakan bahwa model mengenali pola data latih dengan sangat baik.

Tabel 2. Hasil *confusion matrix* data training

		Kelas Variabel	
		Nonzonasi (0)	Zonasi (1)
<i>Actual</i>	Nonzonasi (0)	132	12
	Zonasi (1)	16	704
		<i>Predicted</i>	

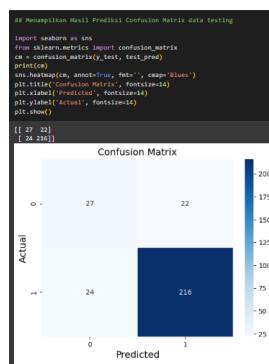


Gambar 8. Hasil *confusion matrix* data training dengan *library seaborn*

Sementara itu, confusion matrix pada data uji (Tabel 3, Gambar 9) menghasilkan akurasi sebesar 84%. Nilai TP tetap dominan, namun jumlah FP dan FN meningkat dibandingkan pada data latih. Perbedaan 13% antara akurasi data latih dan uji mengindikasikan adanya overfitting, yaitu ketika model terlalu menyesuaikan diri dengan data latih sehingga performanya menurun pada data baru.

Tabel 3. Hasil *confusion matrix* data testing

		Kelas variabel	
		Nonzonasi (0)	Zonasi (1)
<i>Actual</i>	Nonzonasi (0)	27	22
	Zonasi (1)	24	216
		<i>Predicted</i>	



Gambar 9. Hasil *confusion matrix* data testing dengan *library seaborn*

Hasil classification report memperkuat temuan ini. Pada data latih (Gambar 10), presisi, recall, dan f1-score masing-masing berada pada kisaran 94–95%, menunjukkan konsistensi model yang sangat baik. Namun, pada data uji (Gambar 11), nilai presisi, recall, dan f1-score turun ke kisaran 72–73%. Hal ini mengindikasikan bahwa meskipun model cukup akurat secara keseluruhan, sensitivitas dan ketepatan dalam mengenali siswa zonasi maupun nonzonasi menurun ketika diuji pada data baru.

## Menampilkan hasil klasifikasi dan tingkat akurasi data training				
from sklearn.metrics import classification_report, accuracy_score				
cr = classification_report(y_train, train_pred)				
ac = accuracy_score(y_train, train_pred)				
print(cr)				
print('Tingkat Akurasi : {:.2f}%'.format(ac*100))				
		precision	recall	f1-score
0		0.89	0.92	0.90
1		0.98	0.98	0.98
				0.97
		accuracy		864
		macro avg	0.94	0.95
		weighted avg	0.97	0.97
Tingkat Akurasi : 96.76%				

Gambar 10. Hasil *classification report* data *training*

## Menampilkan hasil klasifikasi dan tingkat akurasi data testing				
from sklearn.metrics import classification_report, accuracy_score				
cr = classification_report(y_test, test_pred)				
ac = accuracy_score(y_test, test_pred)				
print(cr)				
print('Tingkat Akurasi : {:.2f}%'.format(ac*100))				
		precision	recall	f1-score
0		0.53	0.55	0.54
1		0.91	0.90	0.90
				0.84
		accuracy		289
		macro avg	0.72	0.73
		weighted avg	0.84	0.84
Tingkat Akurasi : 84.08%				

Gambar 11. Hasil *classification report* data *testing*

Perbedaan hasil antara data latih dan uji perlu dicermati dalam konteks implementasi. Dengan akurasi 84%, model cukup handal untuk mendukung proses verifikasi zonasi, tetapi tingkat kesalahan sekitar 16% tidak bisa diabaikan. Oleh karena itu, sistem berbasis KNN lebih tepat digunakan sebagai *decision support system* yang mendampingi keputusan panitia PPDB, bukan sebagai satu-satunya dasar penerimaan siswa

4. KESIMPULAN

Penelitian ini membuktikan bahwa algoritma K-Nearest Neighbors (KNN) efektif digunakan dalam klasifikasi penerimaan peserta didik baru berbasis sistem zonasi di SMPN 16 Bogor, dengan akurasi data latih mencapai 97% dan akurasi data uji sebesar 84%. Hasil ini menunjukkan bahwa atribut spasial berupa koordinat longitude, latitude, serta jarak rumah ke sekolah dapat menjadi dasar yang kuat dalam mendukung proses PPDB. Meskipun demikian, adanya perbedaan signifikan antara akurasi data latih dan data uji mengindikasikan potensi *overfitting*, di mana model terlalu menyesuaikan diri dengan data latih sehingga kinerjanya menurun pada data baru. Kondisi ini menjadi keterbatasan penelitian yang perlu dicermati karena dapat memengaruhi tingkat keandalan sistem ketika diterapkan secara nyata. Oleh karena itu, meskipun model KNN cukup akurat dan layak digunakan sebagai *decision support system* dalam proses seleksi zonasi, sistem ini sebaiknya tidak dijadikan satu-satunya dasar pengambilan keputusan. Penelitian mendatang dapat mengatasi keterbatasan ini dengan memperluas dataset, menerapkan validasi silang (*cross-validation*), atau mengeksplorasi algoritma lain seperti Support Vector Machine (SVM) maupun Naive Bayes untuk meningkatkan kemampuan generalisasi.

REFERENCES

- [1] I. G. P. Darya, “Evaluasi implementasi sistem penerimaan peserta didik baru (PPDB) di Kota Balikpapan, Indonesia,” *J. Penelit. Pendidik.*, vol. 20, no. 1, pp. 32–41, 2020.
- [2] A. Muhammin, “Klasifikasi prestasi akademik siswa berdasarkan nilai rapor dan kedisiplinan.” Universitas Islam Negeri Maulana Malik Ibrahim, 2024.
- [3] V. Chugani, “Understanding Euclidean Distance: From Theory to Practice,” Datacamp. [Online]. Available: <https://www.datacamp.com/tutorial/euclidean-distance>
- [4] M. D. Nawar, F. Helmiah, and C. Latiffani, “Penerapan Metode Weighted Product (Wp) Penilaian Soft Skill Guru Dalam Belajar Di Mts. Hidayatul Ulumiyah Ujung Kubu,” *JIPI (Jurnal Ilm. Penelit. dan Pembelajaran Inform.*, vol. 10, no. 2, pp. 1022–1030, 2025.

- [5] R. K. Imani, S. H. Wijoyo, and F. Amalia, "Penerapan Algoritma K-Nearest Neighbor untuk Klasifikasi Kemampuan Lulusan Siswa Dalam Bersaing untuk Mendapatkan Pekerjaan (Studi Kasus: SMK 'SORE' Tulungagung)," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 8, no. 10, 2024.
- [6] and S. S. J. I. Kartika, E. Santoso, "Penentuan Siswa Berprestasi Menggunakan KNN dan Weighted Product," *J. Pengembang. Tekno. Inf.*, vol. 1, p. no. 5, pp. 352–360, 2017.
- [7] S. Widaningsih, "Penerapan Data Mining untuk Memprediksi Siswa Berprestasi dengan Menggunakan Algoritma K Nearest Neighbor," *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 9, no. 3, pp. 2598–2611, 2022.
- [8] J. A. Samudra, S. Anraeni, and H. Herman, "Penerapan Metode K-Nearest Neighbor Untuk Memprediksi Tingkat Kelulusan Mahasiswa Berbasis Web Pada Fakultas Ilmu Komputer UMI," *Bul. Sist. Inf. dan Teknol. Islam*, vol. 1, no. 4, pp. 230–237, 2020.
- [9] N. Wati, "Prediksi Kelulusan Mahasiswa Menggunakan K-Nearest Neighbor Berbasis Particle Swarm optimization," *J. Teknol. Inf. Indones.*, vol. 6, no. 2, pp. 118–127, 2021.
- [10] A. Putri *et al.*, "Komparasi Algoritma K-NN, Naive Bayes dan SVM untuk Prediksi Kelulusan Mahasiswa Tingkat Akhir: Comparison of K-NN, Naive Bayes and SVM Algorithms for Final-Year Student Graduation Prediction," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 3, no. 1, pp. 20–26, 2023.
- [11] T. Nasution, "Implementasi algoritma K-Nearest Neighbor untuk penentuan kelulusan mahasiswa tepat waktu," *J. Perangkat Lunak*, vol. 2, no. 1, pp. 1–14, 2020.
- [12] H. Henderi, T. Wahyuningsih, and E. Rahwanto, "Comparison of Min-Max normalization and Z-Score Normalization in the K-nearest neighbor (kNN) Algorithm to Test the Accuracy of Types of Breast Cancer," *Int. J. Informatics Inf. Syst.*, vol. 4, no. 1, pp. 13–20, 2021.
- [13] J. Opitz, "A closer look at classification evaluation metrics and a critical reflection of common evaluation practice," *Trans. Assoc. Comput. Linguist.*, vol. 12, pp. 820–836, 2024.
- [14] A. Amelia, M. Asfi, and R. Fahrudin, "Implementation of K-Nearest Neighbor Method for Selection of New Employee Candidates (Case Study: CV. Syntax Corporation Indonesia)," *Eduvest - Journal of Universal Studies*, vol. 4, no. 7, pp. 5742–5754, 2024, doi: 10.5918/eduvest.v4i7.1305.
- [15] 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.
- [16] K. Harian, "Pengertian Dataset dan Jenis-jenisnya," *Kumparan*, 2021. [Online]. Available: <https://kumparan.com/kabar-harian/pengertian-dataset-dan-jenis-jenisnya-1wtM6xNlkpQ>
- [17] V. A. D. Hidayatullah, A. Nilogiri, and H. A. Al Faruq, "Klasifikasi Siswa Berprestasi Menggunakan Metode K-Nearest Neighbor (KNN) Pada SMA Negeri 2 Situbondo," *Jurnal Smart Teknologi*, vol. 1, no. 1, 2021. [Online]. Available: <http://jurnal.unmuhember.ac.id/index.php/JST>
- [18] D. Intern, "Python: Pengertian, Contoh Penggunaan, dan Manfaat Mempelajarinya," *Dicoding Blog*, 2023. [Online]. Available: <https://www.dicoding.com/blog/python-pengertian-contoh-penggunaan-dan-manfaat-mempelajarinya/>
- [19] N. A. Makarim, "Peraturan Menteri Pendidikan, Kebudayaan, Riset, dan Teknologi Republik Indonesia Nomor 17 Tahun 2024 tentang Sistem Zonasi Cagar Budaya," pp. 4–6, Feb. 2024.
- [20] M. Miftah and S. Syamsurijal, "Strategi Pemanfaatan Lingkungan Pendidikan untuk Meningkatkan Motivasi Belajar Siswa," *Edu Cendikia: Jurnal Ilmiah Kependidikan*, vol. 3, no. 01, pp. 72–83, 2023, doi: 10.47709/educendikia.v3i01.2251.
- [21] S. B. Munthe, "Implementasi Algoritma K-Nearest Neighbor Tawar Menggunakan Ekstraksi Fitur ORB," Skripsi, Universitas Medan Area, 2023.
- [22] A. Oliver, "Mengenal Google Colab: Mulai dari Definisi, Cara Menggunakan, hingga Manfaatnya," *Glints*, 2025. [Online]. Available: <https://glints.com/id/lowongan/google-colab-adalah/>
- [23] D. Olivia, "Penerapan Algoritma K-Nearest Neighbor (KNN) untuk Ketepatan Waktu Lulus Mahasiswa," 2024.
- [24] Kumparan, "Pengertian dan Istilah. Apa Itu Zonasi Sekolah? Ini Pengertian, Kelebihan, dan Kekurangannya," *Kumparan*, 2024. [Online]. Available: <https://kumparan.com/pengertian-dan-istilah/apa-itu-zonasi-sekolah-ini-pengertian-kelebihan-dan-kekurangannya-23yziXZJUAH>
- [25] N. Pradita, "Penerapan Data Mining sebagai Cara untuk Memprediksi Prestasi Siswa Berdasarkan Status Ekonomi dan Kedisiplinan Menggunakan Metode Regresi Linier Berganda," *Jurnal Ilmiah Mahasiswa Manajemen, Bisnis dan Akuntansi (JIMMBA)*, vol. 4, no. 5, pp. 683–696, 2022, doi: 10.32639/jimmiba.v4i5.181.
- [26] M. F. Shodiq and D. Darmawan, "Pengaruh Lingkungan Sekolah terhadap Hasil Belajar Siswa Sekolah Menengah Pertama," *IHSANIKA: Jurnal Pendidikan Agama Islam*, vol. 3, no. 1, pp. 292–307, 2025.
- [27] R. S. Wahono, "Data Mining," in *Mining of Massive Datasets*, vol. 2, Jan. 2013. [Online]. Available: https://www.cambridge.org/core/product/identifier/CBO9781139058452A007/type/book_part
- [28] Y. Yahya and W. Hidayanti, "Penerapan Algoritma K-Nearest Neighbor Untuk Klasifikasi Efektivitas Penjualan Vape (Rokok Elektrik) pada 'Lombok Vape On,'" *Infotek: Jurnal Informatika dan Teknologi*, vol. 3, pp. 104–114, 2020, doi: 10.29408/jit.v3i2.2279.