



Pemodelan K-Nearest Neighbor Untuk Identifikasi Pola Kepuasan Mahasiswa Terhadap Pelayanan Kampus (Studi Kasus : STMIK Kaputama)

Muhammad Rizky R Ritonga^{1*}, Marto Sihombing², Selfira³

^{1,2,3} STMIK Kaputama, Indonesia

rizkyritonga07@gmail.com^{1*}, martosihombing45@gmail.com², selfira.yap@gmail.com³

Alamat: Jl. Veteran No.4A, Tangsi, Kec. Binjai Kota, Kota Binjai, Sumatera Utara 20714

Korespondensi penulis: rizkyritonga07@gmail.com

Abstract. This research focuses on using the K-Nearest Neighbor (KNN) algorithm to model student satisfaction with campus services. The study finds that the quality of the dataset strongly influences the accuracy of the KNN classification results. Factors such as data cleanliness, balanced class distribution, and sufficient training data volume are highlighted as crucial for a successful model. The research also emphasizes the significance of proper feature selection in enhancing classification performance, suggesting that irrelevant features can introduce noise and decrease model accuracy. The model was evaluated using a dataset of 1032 data points and $K=5$, achieving an accuracy of 93.72%. While the model performed well for certain classes such as "Very Good" and "None", challenges were encountered in classifying the "Fair" and "Deficient" classes. The study concludes that KNN is effective in identifying student satisfaction patterns but highlights the need for improvements in accurately classifying these challenging classes. Ultimately, the research underscores the importance of data quality and feature selection in enhancing the performance of classification models for student satisfaction analysis.

Keywords: Classification, KNN, Satisfaction, Services, Student

Abstrak: Penelitian ini berfokus pada penggunaan algoritma K-Nearest Neighbor (KNN) untuk memodelkan kepuasan mahasiswa terhadap layanan kampus. Penelitian ini menemukan bahwa kualitas dataset sangat mempengaruhi akurasi hasil klasifikasi KNN. Faktor-faktor seperti kebersihan data, distribusi kelas yang seimbang, dan volume data pelatihan yang cukup disoroti sebagai hal yang penting untuk model yang sukses. Penelitian ini juga menekankan pentingnya pemilihan fitur yang tepat dalam meningkatkan kinerja klasifikasi, menunjukkan bahwa fitur yang tidak relevan dapat menimbulkan noise dan menurunkan akurasi model. Model tersebut dievaluasi dengan menggunakan dataset yang terdiri dari 1032 titik data dan $K=5$, mencapai akurasi 93,72%. Meskipun model tersebut berkinerja baik untuk kelas-kelas tertentu seperti "Sangat Baik" dan "Tidak Ada", ada beberapa tantangan yang dihadapi dalam mengklasifikasikan kelas "Cukup" dan "Kurang". Penelitian ini menyimpulkan bahwa KNN efektif dalam mengidentifikasi pola kepuasan siswa, tetapi menyoroti perlunya perbaikan dalam mengklasifikasikan kelas-kelas yang menantang ini secara akurat. Pada akhirnya, penelitian ini menggarisbawahi pentingnya kualitas data dan pemilihan fitur dalam meningkatkan kinerja model klasifikasi untuk analisis kepuasan siswa.

Kata kunci: Klasifikasi, KNN, Kepuasan, Layanan, Mahasiswa

1. PENDAHULUAN

Kepuasan mahasiswa terhadap pelayanan kampus menjadi faktor kunci dalam menentukan kualitas pendidikan dan daya saing suatu institusi. Kampus yang mampu memberikan pelayanan prima akan lebih diminati oleh mahasiswa dan memiliki peluang berkembang lebih pesat. STMIK KAPUTAMA, sebagai salah satu institusi pendidikan tinggi, berkomitmen untuk terus meningkatkan kualitas layanan melalui survei kepuasan mahasiswa.

Survei kepuasan mahasiswa memberikan informasi penting mengenai persepsi mahasiswa terhadap kualitas layanan kampus. Data ini dapat digunakan untuk mengidentifikasi area yang perlu ditingkatkan serta mengembangkan strategi pelayanan yang lebih baik. Salah satu metode yang dapat digunakan untuk menganalisis hasil survei ini adalah algoritma K-Nearest Neighbors (KNN).

KNN bekerja dengan mencari tetangga terdekat dari data baru berdasarkan data latih yang ada. Metode ini memprediksi kelas dari data baru dengan mempertimbangkan mayoritas kelas dari K tetangga terdekat. Meskipun sederhana dan mudah diimplementasikan, KNN sensitif terhadap pemilihan nilai K dan kinerjanya dapat terpengaruh oleh data berdimensi tinggi.

Data mining berperan penting dalam menemukan pola tersembunyi dari data yang besar. Penggunaan data mining dengan metode KNN memungkinkan analisis yang lebih mendalam untuk mengidentifikasi pola kepuasan mahasiswa terhadap pelayanan kampus, meskipun data awal seringkali masih mentah dan memerlukan pengolahan lebih lanjut agar siap digunakan dalam penelitian.

2. METODE PENELITIAN

Penelitian Terdahulu

Berdasarkan penelitian sebelumnya oleh Dwi Fasnuari et al. (2022), metode K-Nearest Neighbor (KNN) menghasilkan akurasi sebesar 67%, dengan presisi 65%, recall 73%, dan f-measure 96% pada nilai $K=250$ menggunakan jarak Manhattan. Sementara itu, metode Naïve Bayes Classifier hanya mencapai akurasi 58%, dengan presisi 90%, recall 55%, dan f-measure 68%. Dari hasil ini, dapat disimpulkan bahwa metode KNN memiliki performa yang lebih baik dalam klasifikasi dataset penyakit jantung.

Selain itu, penelitian oleh (Malik Namus Akbar, n.d.) yang menggunakan data dari 200 sampel air sumur di Kota Surakarta menunjukkan bahwa algoritma KNN mampu mengidentifikasi kualitas air dengan akurasi 82,6%. Hasil ini menunjukkan bahwa metode KNN efektif digunakan untuk mengklasifikasikan kualitas air minum, dengan tingkat akurasi yang tinggi, yaitu hingga 82%.

Berdasarkan hasil penelitian yang telah dilakukan sebelumnya dengan hasil yang diperoleh dari hasil prediksi yang menunjukkan bahwa prediksi menggunakan metode K-Nearest Neighbor berhasil ditetapkan, didapatkan dari 25 data testing memperoleh hasil dengan nilai $k = 7$ memiliki keputusan yaitu terlaris = 4 dan tidak laris = 3. Dari penelitian tersebut bisa menjadi salah satu acuan dalam tugas akhir ini untuk menentukan penjualan

sepeda motor masuk kedalam kategori laris ataupun tidak laris. Selain itu dilihat dari penelitian tersebut menunjukkan bahwa pemilihan algoritma K-NN ini menjadi salah satu cara yang paling mudah untuk digunakan dalam tugas akhir ini mengenai penentuan label laris dan tidak laris pada penjualan sepeda motor di PT. Sumber Rejeki Jabar Cirebon. (Ali & Rizki Rinaldi, 2023)

Teori Pendukung Penelitian

a. Data Mining

(Mai et al., 2022) Data mining merupakan serangkaian proses untuk menggali nilai tambah berupa informasi yang selama ini tidak diketahui secara manual dari suatu basis data. Data mining mulai ada sejak 1990-an sebagai cara yang benar dan tepat untuk mengambil pola dan informasi yang digunakan untuk menemukan hubungan antara data untuk melakukan pengelompokan ke dalam satu atau lebih cluster sehingga objek-objek yang berada dalam satu cluster akan mempunyai kesamaan yang tinggi antara satu dengan lainnya. Data mining merupakan bagian dari proses penemuan pengetahuan dari basis data Knowledge Discovery in Databases.

b. K-Nearest Neighbor

Menurut (Cholil et al., 2021) K-Nearest Neighbor (K-NN) merupakan salah satu algoritma terpandu di mana hasil klasifikasi dari instance baru ditentukan berdasarkan mayoritas kategori tetangga terdekat. K-NN mengklasifikasikan data berdasarkan kedekatan atau jaraknya dengan data lain. Algoritma KNN dianggap sebagai metode non-parametrik berbasis contoh yang paling sederhana dalam proses data mining.

Meskipun KNN mudah diimplementasikan pada dataset berukuran kecil, metode ini memiliki keterbatasan jika diterapkan pada dataset yang besar dan kompleks, karena waktu prosesnya menjadi tidak efisien. KNN mengklasifikasikan data baru dengan menghitung jarak ke K tetangga terdekat, dan label kelas dari tetangga tersebut digunakan untuk memprediksi kelas instance baru. Pemilihan jumlah K yang tepat sangat penting karena K yang terlalu kecil membuat algoritma sensitif terhadap noise, sedangkan K yang terlalu besar dapat menyebabkan bias dalam model.

Sebagai algoritma berbasis memori, KNN melakukan iterasi pada data hingga menemukan atribut atau parameter yang paling dekat. Jarak minimum

antara data uji dan data latih dibandingkan untuk menentukan klasifikasi. Beberapa metode pengukuran jarak yang digunakan dalam KNN adalah jarak Manhattan (city block distance) dan jarak Euclidean, dengan jarak Euclidean menjadi metode yang paling umum digunakan. Pengukuran jarak yang paling sering digunakan adalah *euclidean distance*. Jarak *euclidean distance* didefinisikan seperti pada persamaan :

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

Keterangan :

d = Jarak

i = Variabel data

p = Dimensi data

X1 = Data Training

X2 = Data Testing

c. Confusion Matrix

Menurut (Dwi Fasnuari et al., 2022), confusion matrix digunakan untuk membandingkan nilai asli dengan nilai prediksi model, sehingga dapat menilai kinerja model secara menyeluruh. Confusion matrix merupakan alat evaluasi yang berguna dalam model klasifikasi, karena memungkinkan pengukuran performa dengan cara mengidentifikasi sejauh mana hasil prediksi sesuai dengan label yang sebenarnya. Matriks ini menyajikan jumlah prediksi yang benar dan salah dalam empat kategori:

- a) True Positives (TP): Jumlah kasus di mana model dengan benar memprediksi kelas positif.
- b) True Negatives (TN): Jumlah kasus di mana model dengan benar memprediksi kelas negatif.
- c) False Positives (FP): Jumlah kasus di mana model salah memprediksi kelas positif (type I error).
- d) False Negatives (FN): Jumlah kasus di mana model salah memprediksi kelas negatif (type II error).

Confusion matriks memiliki 3 rumus yaitu :

a) Presisi

Presisi merupakan tingkat keberhasilan model dalam memberikan jawaban dengan tepat kepada pengguna.

$$\text{Presisi} = \frac{Tp}{TP+FP}$$

b) Recall

Recall merupakan tingkat keberhasilan model dalam menemukan kembali informasi dengan benar.

$$\text{Recal} = \frac{Tp}{TP+FN}$$

c) F1-score

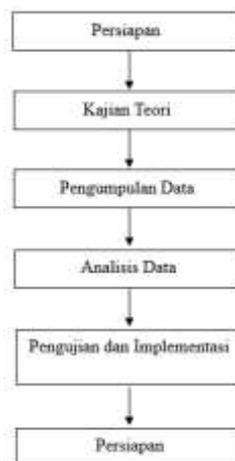
F1-Score merupakan hasil perbandingan antara nilai presisi dan *recall*. F1-score adalah sebuah metrik yang digunakan untuk mengukur kinerja model klasifikasi.

$$\text{F1-Score} = \frac{2 \times \text{presisi} \times \text{recall}}{\text{presisi} + \text{recall}}$$

3. ANALISIS DAN PERANCANGAN

a. Metode penelitian

Metode penelitian ini dilakukan secara sistematis menggunakan metode ilmiah dan sumber-sumber yang relevan. Proses ini bertujuan untuk menghasilkan penelitian yang akurat dan tepat. Berdasarkan metode penelitian yang digunakan, dapat disusun suatu alur kegiatan yang jelas dan terstruktur untuk mencapai hasil yang optimal.



Gambar 1 Metode penelitian

1. Persiapan: Menentukan latar belakang masalah, batasan, tujuan, dan manfaat untuk menyusun proses kerja sistem.

2. Kajian Teori: Mengumpulkan teori dari buku, jurnal, dan internet yang mendukung penelitian, termasuk topik tentang Data Mining, K-Nearest Neighbor, dan kepuasan mahasiswa.
3. Pengumpulan Data: Mengumpulkan data yang diperlukan dari STMIK KAPUTAMA untuk penelitian.
4. Analisa Data: Mengelola dan menganalisis data yang telah dikumpulkan, melakukan transformasi data, dan memprosesnya dengan metode K-Nearest Neighbor.
5. Pengujian dan Implementasi: Melakukan validasi, implementasi, dan pengujian data menggunakan Software Rapidminer.
6. Akhir: Menyusun kesimpulan dan saran berdasarkan hasil penelitian untuk perbaikan di masa depan.

b. Skema Proses K-Nearest Neighbor



Gambar 2 Skema KNN

Pada flowchart tersebut, langkah pertama adalah memasukkan data training. Setelah itu, nilai K ditentukan sebagai jumlah tetangga terdekat yang akan digunakan dalam klasifikasi. Kemudian, dilakukan perhitungan jarak Euclidean antara data training dan data yang akan diklasifikasikan. Jarak tersebut diurutkan, dan hasil klasifikasi ditentukan berdasarkan nilai yang paling sering muncul di antara tetangga terdekat. Setelah proses klasifikasi selesai, langkah terakhir adalah melakukan evaluasi model untuk mengukur akurasi.

c. Data pendukung

Untuk menganalisis data dalam penelitian, diperlukan data pendukung agar penelitian dapat berjalan sesuai harapan. Berdasarkan penelitian yang dilakukan di STMIK KAPUTAMA, data yang diperoleh akan digunakan untuk menganalisis kepuasan mahasiswa terhadap pelayanan kampus. Data tersebut menjadi dasar dalam proses analisis untuk mendapatkan hasil yang relevan dan bermanfaat bagi penelitian.

Tabel 1 keterangan aspek pelayanan kampus

No	Nama	Keterangan
1	Aspek 1	Kenyamanan ruang kuliah teori
2	Aspek 2	Kenyamanan dan ketersediaan Laboratorium 1
3	Aspek 3	Kenyamanan dan ketersediaan Laboratorium 2
4	Aspek 4	Kenyamanan dan ketersediaan Laboratorium 3
5	Aspek 5	Kelengkapan alat praktik untuk praktikum
6	Aspek 6	Pelayanan bagian laboratorium (Asisten LAB dan Kepala LAB)
7	Aspek 7	Pelayanan bagian perpustakaan
8	Aspek 8	Jumlah dan mutu buku, diktat, jurnal, dll, di perpustakaan
9	Aspek 9	Pelayanan bagian keuangan
10	Aspek 10	Pelayanan administrasi akademik
11	Aspek 11	Pelayanan bimbingan akademik (Dosen penasehat akademik)
12	Aspek 12	Pelayanan keamanan kampus (Satpam, parkir kendaraan, dll,)
13	Aspek 13	Kebersihan lingkungan kampus (Taman, kamar mandi, dll,)
14	Aspek 14	Kenyamanan dan ketersediaan sarana ibadah (musholla, tempat wudhu, dll,)
15	Aspek 15	Ketersediaan sarana olahraga (Lapangan, peralatan, dll,)
16	Aspek 16	Cakupan dan bandwidth hotspot (wifi) kampus
17	Aspek 17	Pelayanan sistem informasi (website, pengumuman elektronik, dll,)

Tabel 2 Dataset Kepuasan Mahasiswa Terhadap Pelayanan Kampus

No	Nama	Aspek 1	Aspek 2	Aspek 3	Aspek ...	Aspek 15	Aspek 16	Aspek 17	Status
1	A	5	4	4	...	4	2	2	Cukup
2	B	4	3	3	...	3	3	4	Cukup
3	C	4	4	4	...	4	4	4	Baik
4	D	5	4	4	...	4	4	4	Baik
5	E	5	4	5	...	3	5	4	Baik
6	F	3	3	3	...	3	4	4	Cukup
7	G	4	4	4	...	4	4	4	Cukup
8	H	4	4	4	...	2	4	4	Cukup

9	I	4	4	4	...	3	4	4	Cukup
10	J	4	4	4	...	4	3	4	Cukup
11	K	4	4	4	...	4	4	4	Baik
12	L	4	4	4	...	4	4	4	Baik
13	M	3	3	4	...	2	3	4	Cukup
14	N	2	3	3	...	2	2	4	Kurang
15	O	5	4	4	...	4	4	5	Baik
16	P	4	4	4	...	4	4	4	Baik
17	Q	5	5	4	...	5	5	5	Baik
18	R	3	1	2	...	1	1	2	Kurang
19	S	4	4	4	...	3	4	3	Cukup
20	T	3	2	4	...	4	2	3	Cukup
21	U	2	1	1	...	1	1	2	?

Tabel 3 Interval Nilai Kepuasan Mahasiswa

Kategori	Interval
Tidak ada	0
Sangat kurang	1 – 1,9
Kurang	2 – 2,9
Cukup	3 – 3,9
Baik	4 – 4,9
Sangat baik	5

d. Penerapan KNN

Setelah data training mahasiswa ditentukan, langkah selanjutnya adalah melakukan proses klasifikasi pada data testing untuk mahasiswa ke-21. Tujuan dari proses ini adalah untuk menentukan apakah tingkat kepuasan mahasiswa tersebut termasuk dalam kategori Baik, Cukup, atau Kurang. Proses klasifikasi K-Nearest Neighbor ini akan dilakukan dengan langkah-langkah berikut:

1. Menentukan parameter K, di dalam penelitian ini parameter K = 5.
2. Menghitung jarak data training dan data testing menggunakan rumus Euclidean Distance

Tabel 4 Perhitungan Euclidean Distance Pada Mahasiswa

No	Euclidean Distance
1	$= \sqrt{(5-2)^2 + (4-1)^2 + (4-1)^2 + (4-1)^2 + (4-1)^2 + (4-2)^2 + (4-2)^2 + (4-2)^2 + (4-1)^2 + (4-2)^2 + (4-2)^2 + (4-2)^2 + (4-2)^2 + (4-2)^2 + (4-1)^2 + (2-1)^2 + (2-2)^2}$ $= \sqrt{9+9+9+9+9+4+4+4+9+4+4+4+4+4+9+1+0}$ $= \sqrt{96}$ $= 9,797958971 \text{ Jarak A}$

2	$= \sqrt{(4-2)^2 + (3-1)^2 + (3-1)^2 + (4-1)^2 + (3-1)^2 + (3-2)^2 + (3-2)^2 + (4-2)^2 + (4-1)^2 + (4-2)^2 + (4-2)^2 + (4-2)^2 + (4-2)^2 + (4-2)^2 + (3-1)^2 + (3-1)^2 + (4-2)^2}$ $= \sqrt{4 + 4 + 4 + 9 + 4 + 1 + 1 + 4 + 9 + 4 + 4 + 4 + 4 + 4 + 4 + 4 + 4 + 4 + 4}$ $= \sqrt{72}$ $= 8,485281374 \text{ Jarak B}$
...	...
19	$= \sqrt{(4-2)^2 + (4-1)^2 + (4-1)^2 + (4-1)^2 + (2-1)^2 + (4-2)^2 + (4-2)^2 + (4-2)^2 + (3-1)^2 + (3-2)^2 + (3-2)^2 + (4-2)^2 + (4-2)^2 + (4-2)^2 + (3-1)^2 + (4-1)^2 + (3-2)^2}$ $= \sqrt{4 + 9 + 9 + 9 + 1 + 4 + 4 + 4 + 4 + 1 + 1 + 4 + 4 + 4 + 4 + 9 + 1}$ $= \sqrt{76}$ $= 8,717797887 \text{ Jarak S}$
20	$= \sqrt{(3-2)^2 + (2-1)^2 + (4-1)^2 + (3-1)^2 + (4-1)^2 + (3-2)^2 + (3-2)^2 + (4-2)^2 + (3-1)^2 + (3-2)^2 + (3-2)^2 + (4-2)^2 + (5-2)^2 + (4-2)^2 + (4-1)^2 + (2-1)^2 + (3-2)^2}$ $= \sqrt{1 + 1 + 9 + 4 + 9 + 1 + 1 + 4 + 4 + 1 + 1 + 4 + 9 + 4 + 9 + 1 + 1}$ $= \sqrt{64}$ $= 8 \text{ Jarak T}$

e. Pengukuran Confusion Matrix

Setelah dilakukan prediksi kelas terhadap data uji langkah terakhir yang perlu dilakukan adalah mengukur keakuratan model dalam melakukan prediksi. Berikut adalah hasil perhitungan akurasi dengan menggunakan metrik confusion matrix dengan menggunakan dua puluh data latih dengan dua puluh data uji yang di asumsikan telah memiliki label hasil prediksi :

Nama	Aspek 1 S/d 17	Label Sebenarnya	Prediksi
A	...	Cukup	Baik
B	...	Cukup	Kurang
C	...	Baik	Cukup
D	...	Baik	Baik
E	...	Baik	Baik
F	...	Cukup	Cukup
G	...	Cukup	Baik
H	...	Cukup	Baik
I	...	Cukup	Cukup
J	...	Cukup	Baik
K	...	Baik	Baik
L	...	Baik	Baik
M	...	Cukup	Baik
N	...	Kurang	Baik
O	...	Baik	Baik
P	...	Baik	Baik
Q	...	Baik	Kurang
R	...	Kurang	Cukup
S	...	Cukup	Cukup
T	...	Cukup	Cukup

Gambar 3 Sample Data Hasil Prediksi

Prediksi	Kurang	Cukup	Baik
True Positive	0	4	6
False Positive	2	2	6
False Negative	2	6	2

Gambar 4 Sample Data Hasil Prediksi

Dari hasil interpretasi di atas dapat dilakukan perhitungan confusion matrix :

$$\begin{array}{l}
 \text{Presisi Kurang} \quad \text{Recall Kurang} \quad \text{F1-Score kurang} \\
 = \frac{0}{0+2} = 0 \quad = \frac{0}{0+2} = 0 \quad = \frac{2 \times 0 \times 0}{0+0} = 0 \\
 \\
 \text{Presisi cukup} \quad \text{Recall cukup} \quad \text{F1-score cukup} \\
 = \frac{4}{4+2} = \frac{4}{6} = 0,6 \quad = \frac{4}{4+6} = \frac{4}{10} = 0,4 \quad = \frac{2 \times 0,6 \times 0,4}{0,6+0,4} = \frac{0,48}{1} = 0,48 \\
 \\
 \text{Presisi Baik} \quad \text{Recall Baik} \quad \text{F1-Score baik} \\
 = \frac{6}{6+6} = \frac{6}{12} = 0,5 \quad = \frac{6}{6+2} = \frac{6}{8} = 0,75 \quad = \frac{2 \times 0,5 \times 0,75}{0,5+0,75} = \frac{0,75}{1,25} = 0,6
 \end{array}$$

4. KESIMPULAN

Penelitian ini menunjukkan bahwa kualitas dataset sangat mempengaruhi hasil klasifikasi KNN, di mana data yang bersih, seimbang, dan relevan meningkatkan performa, sedangkan noise dan ketidakseimbangan kelas menurunkannya. Relevansi fitur juga penting, karena fitur yang signifikan lebih memengaruhi hasil klasifikasi. Pengujian KNN dengan 1032 data dan K=5 menghasilkan akurasi 93.72%, dengan recall tinggi untuk beberapa kelas, meskipun kelas Cukup dan Kurang menunjukkan kesulitan dalam klasifikasi.

DAFTAR REFERENSI

- Ali, I., & Rizki Rinaldi, A. (2023). Penerapan metode k-nearest neighbor untuk prediksi penjualan sepeda motor terlaris. *Jurnal Mahasiswa Teknik Informatika*, 7(1).
- Cholil, S. R., Handayani, T., Prathivi, R., & Ardianita, T. (2021). Implementasi algoritma klasifikasi k-nearest neighbor (KNN) untuk klasifikasi seleksi penerima beasiswa. *IJCIT (Indonesian Journal on Computer and Information Technology)*, 6(2).
- Dwi Fasnuari, H. A., Yuana, H., & Chulkamdi, M. T. (2022). Penerapan algoritma k-nearest neighbor untuk klasifikasi penyakit diabetes melitus. *Antivirus: Jurnal Ilmiah Teknik Informatika*, 16(2), 133–142. <https://doi.org/10.35457/antivirus.v16i2.2445>
- Mai, P., Tarigan, S., Tata Hardinata, J., Qurniawan, H., Safii, M., Winanjaya, R., Studi, P., Informasi, S., Tunas, S., & Pematangsiantar, B. (2022). Implementasi data mining menggunakan algoritma apriori dalam menentukan persediaan barang (studi kasus: Toko Sinar Harahap). *Jurnal Umum Manajemen dan Teknologi Informasi*, 12(2). <https://jurnal.umj.ac.id/index.php/just-it/index>
- Malik Namus Akbar, F. (n.d.). Metode KNN (k-nearest neighbor) untuk menentukan kualitas air. *I8(1)*.