

PERBANDINGAN METODE NAÏVE BAYES DAN K-NEAREST NEIGHBORS DALAM KLASIFIKASI KEPUASAN MAHASISWA TERHADAP LAYANAN WI-FI DI UNIVERSITAS NASIONAL

Fadli Aufa Hawari, Ira Diana Sholihati
Sistem Informasi, Universitas Nasional
Jl. Sawo Manila No.61 Jakarta, Indonesia
fadliaufahawari2021@student.unas.ac.id

ABSTRAK

Layanan Wi-Fi di perguruan tinggi sangat penting bagi mahasiswa untuk menunjang kegiatan akademiknya. Namun, kualitas layanan Wi-Fi di Universitas Nasional masih menjadi perhatian utama karena berbagai keluhan pengguna. Penelitian ini bertujuan untuk membandingkan kinerja algoritma Naïve Bayes dan K-Nearest Neighbor (KNN) dalam mengklasifikasi tingkat kepuasan mahasiswa terhadap layanan Wi-Fi. Data dikumpulkan melalui survei berbasis Google Forms dan dianalisis menggunakan Google Colab. Evaluasi kinerja dilakukan berdasarkan metrik accuracy, precision, recall, dan F1-score. Hasil penelitian menunjukkan bahwa kinerja metode Naïve Bayes dalam mengklasifikasi kepuasan mahasiswa mencapai Accuracy 100%, Precision 100%, Recall 100%, dan F1-score 100%. Hal ini lebih tinggi dibandingkan kinerja metode K-Nearest Neighbors yang mana Accuracy mencapai 86%, Precision 89,8%, Recall 87,2%, dan F1-Score 87,6%. Temuan ini dapat menjadi referensi bagi universitas untuk meningkatkan layanan Wi-Fi berdasarkan analisis kepuasan mahasiswa.

Kata kunci : Klasifikasi, Kepuasan Siswa, Naïve Bayes, K-Nearest Neighbors, Evaluasi Kinerja.

1. PENDAHULUAN

Penelitian ini mengacu pada perbandingan kinerja metode klasifikasi K-Nearest Neighbor (KNN) dan Naive Bayes pada klasifikasi dokumen teks [1].

Metode Naïve Bayes merupakan metode klasifikasi probabilistik yang menghitung probabilitas setiap kelas dan menentukan kelas dengan probabilitas tertinggi. Keunggulan metode ini terletak pada kesederhanaan dan efektivitasnya dalam menangani variabel yang diasumsikan independen [2].

Penelitian ini mengusulkan sistem lokalisasi dalam ruangan berbasis sinyal WiFi menggunakan algoritma Cluster K-Nearest Neighbor (KNN) [3].

Pengujian dilakukan dengan menggunakan metode Naïve Bayes dengan berbagai skenario data latih dan data uji [4].

Penelitian Xingbin Ge membahas tentang optimasi algoritma K-Nearest Neighbor (KNN) untuk pemetaan lokasi dalam ruangan menggunakan sinyal WiFi [5].

Hasil ini menunjukkan bahwa algoritma Naïve Bayes efektif dalam mengklasifikasi kepuasan mahasiswa berdasarkan atribut layanan kampus, sehingga dapat digunakan untuk mengevaluasi kualitas layanan lembaga pendidikan di wilayah tersebut [6].

2. TINJAUAN PUSTAKA

2.1. Penelitian Terdahulu

Terdapat beberapa jurnal relevan yang menjadi referensi dalam penelitian ini. Hasil ditemukan yang menggunakan metode Naïve Bayes untuk memprediksi kepuasan layanan akademik di Universitas Peradaban. Dengan menggunakan 100 data latih dan 17 data uji, sistem hanya mencapai akurasi sebesar 52%. Hal ini menunjukkan bahwa

performa Naïve Bayes sangat bergantung pada jumlah data latih yang tersedia [7].

Penelitian oleh Kiswara et al menunjukkan bahwa algoritma C4.5 dapat mengklasifikasikan tingkat kepuasan mahasiswa yang berlangganan WiFi Indihome dengan akurasi mencapai 75%. Data dikumpulkan melalui observasi dan kuesioner kepada pelanggan PT Telkom Indihome di Pematangsiantar. Hasil penelitian menunjukkan bahwa mayoritas mahasiswa merasa puas terhadap layanan yang diberikan. Namun, kualitas jaringan dan kecepatan akses masih menjadi kendala utama yang perlu diperbaiki. Rekomendasi dari penelitian mencakup peningkatan infrastruktur jaringan serta optimasi layanan berbasis data mining untuk meningkatkan pengalaman pelanggan [8].

Penelitian oleh Ima Wardani menunjukkan bahwa harga dan fasilitas memiliki pengaruh signifikan terhadap kepuasan pelanggan di Kedai Coffee JMP Pahlawan Lamongan, sementara layanan WiFi memiliki pengaruh negatif tetapi tidak signifikan. Data diperoleh dari survei terhadap 95 dan dianalisis menggunakan regresi linier berganda. Hasilnya menunjukkan bahwa peningkatan kualitas fasilitas dan penyesuaian harga dapat meningkatkan kepuasan pelanggan secara keseluruhan [9].

2.2. Naïve Bayes

Naïve Bayes adalah metode klasifikasi yang didasarkan pada teorema Bayes dengan asumsi independensi antara fitur-fitur dalam dataset. Metode ini sering digunakan karena kemampuannya yang sederhana namun efektif dalam menganalisis data berlabel [10].

$$P(Y|X) = \frac{P(X|Y) \cdot P(Y)}{P(X)} \quad (1)$$

2.3. K-Nearest Neighbor (KNN)

KNN adalah metode klasifikasi berbasis instance yang menentukan kelas data baru berdasarkan kedekatan jarak dengan sejumlah data tetangga (*nearest neighbor*) [11].

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (2)$$

2.4. Perbandingan Naïve Bayes dan KNN

Penelitian yang membandingkan kedua metode ini menunjukkan bahwa keduanya memiliki kelebihan masing-masing. Naïve Bayes lebih unggul dalam analisis data yang memiliki distribusi probabilitas tertentu, sementara KNN lebih baik pada data yang tidak memiliki asumsi distribusi tertentu, asalkan metrik jarak dan parameter k dipilih secara optimal [12].

2.5. Google Colab

Google Colab adalah platform berbasis cloud yang menyediakan lingkungan pemrograman Python untuk analisis data, pembelajaran mesin, dan eksperimen berbasis data lainnya. Dengan integrasi pustaka populer seperti Pandas, NumPy, Scikit-learn, dan Matplotlib, platform ini memudahkan pemrosesan data, pengembangan model. Dalam penelitian ini, Google Colab digunakan untuk mengimplementasikan algoritma Naïve Bayes dan K-Nearest Neighbor (KNN) untuk klasifikasi tingkat kepuasan mahasiswa terhadap layanan Wi-Fi, di mana pustaka Scikit-learn dipakai untuk membangun model algoritma, sementara Matplotlib dan Seaborn digunakan untuk visualisasi hasil analisis [13].

2.6. Evaluasi Model Klasifikasi

Pengukuran terhadap kinerja suatu proses pengklasifikasian merupakan hal yang penting. Kinerja proses klasifikasi menggambarkan seberapa baik sistem dalam mengklasifikasikan data. Pengukuran kinerja proses tersebut dapat menggunakan confusion matrix. Confusion matrix adalah sebuah matriks yang memuat data klasifikasi yang dilakukan oleh sistem klasifikasi baik secara aktual maupun prediktif. Confusion matrix berbentuk tabel matriks dengan kinerja model klasifikasi pada serangkaian data testing yang nilainya diketahui yang dijelaskan:

Tabel 1. Confision Matrix

Predicted Data	Actual Data		
		positive	Negative
	Positive	(TP)	(FP)
	Negative	(FN)	(TN)

Keterangan:

- True positive (TP), data yang kelas aktualnya adalah kelas (+) terdeteksi benar kelas (+).

- False negative (FN), data yang kelas aktualnya adalah kelas (+) namun terdeteksi kelas (-).
- False positive (FP), data yang kelas aktualnya adalah kelas (-) namun terdeteksi kelas (+).
- True negative (TN), data yang kelas aktualnya adalah kelas (-) terdeteksi benar kelas (-).

3. METODE PENELITIAN

Dalam desain penelitian ini, yang menggunakan pendekatan kuantitatif. Pendekatan kuantitatif dipilih karena data penelitian ini berbentuk numerik, serta untuk memberikan hasil analisis yang objektif dan terukur.

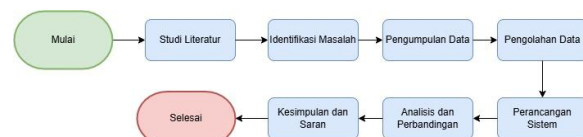
3.1. Teknik Pengumpulan Data

Berikut tahapan pendekatan pengumpulan data yang digunakan dalam penelitian ini:

- Survei: Data dikumpulkan melalui Google Form yang disebarakan kepada mahasiswa Universitas Nasional yang menggunakan layanan Wi-Fi kampus.
- Observasi: Observasi dilakukan untuk mendapatkan wawasan tambahan terkait penggunaan layanan Wi-Fi di Universitas Nasional.
- Studi Pustaka: Dalam langkah ini untuk mendalami konsep-konsep yang relevan dengan penelitian ini seperti penerapan metode klasifikasi dalam analisis data kepuasan.

3.2. Tahap Penelitian

Beberapa tahapan menjadi kerangka penelitian dalam penelitian ini. Gambar 3.1 memberikan rincian lengkap tahapan penelitian.



Gambar 2. Tahapan penelitian

3.3. Perancangan Sistem

Pada bagian ini, akan dijelaskan langkah-langkah yang dilakukan untuk membandingkan kinerja dua metode klasifikasi, yaitu Naïve Bayes dan K-Nearest Neighbors (KNN), dalam mengklasifikasikan tingkat kepuasan mahasiswa terhadap layanan Wi-Fi di Universitas Nasional. Sistem ini akan diimplementasikan menggunakan dataset yang mencakup tanggapan mahasiswa terhadap berbagai aspek layanan Wi-Fi di kampus.

3.3.1. Deskripsi Dataset

Dataset yang digunakan berisi data dari survei kepuasan mahasiswa terkait layanan Wi-Fi di Universitas Nasional.

Tabel 2. Tabel dataset layanan Wi-Fi

No	Pembarayan	Kecepatan	Stabilitas	Kesulitan	Cakupan	Kepuasan
1	Ada	Sangat Baik	Buruk	Ada	Buruk	Sangat Puas
2	Tidak	Sangat Buruk	Baik	Tidak	Baik	Sangat Puas
3	Ada	Baik	Buruk	Tidak	Buruk	Sangat Puas
4	Tidak	Buruk	Baik	Tidak	Baik	Sangat Puas
5	Ada	Buruk	Baik	Ada	Buruk	Tidak Puas

3.3.2. Preprocessing Data

Sebelum menerapkan kedua algoritma, langkah-langkah berikut dilakukan pada dataset:

- Pembersihan Data: Menghapus data yang hilang atau tidak valid.
- Konversi Fitur Kategorikal: Mengonversi variabel kategorikal menjadi variabel numerik menggunakan teknik Label Encoding.
- Pembagian Data: Dataset akan dibagi menjadi dua bagian:
 - Training Set (90% dari total data)
 - Testing Set (10% dari total data)
- Normalisasi Data: Untuk KNN, fitur numerik akan dinormalisasi agar jarak antar data tidak terdistorsi oleh perbedaan skala antar fitur.

3.3.3. Implementasi Model

Setelah data diproses, dua algoritma klasifikasi akan diterapkan untuk memprediksi tingkat kepuasan mahasiswa terhadap layanan Wi-Fi.

Naïve Bayes:

- Model Naïve Bayes diterapkan untuk klasifikasi berdasarkan distribusi probabilitas dari masing-masing fitur. Algoritma ini mengasumsikan bahwa semua fitur bersifat independen, meskipun pada kenyataannya ini mungkin tidak selalu benar.

K-Nearest Neighbors (KNN):

- Model KNN akan digunakan untuk mengklasifikasikan mahasiswa berdasarkan kedekatannya dengan data lain dalam ruang fitur.
- Nilai K (jumlah tetangga terdekat) akan dipilih dengan eksperimen, biasanya diuji dengan nilai ganjil seperti 3, 5, atau 7, agar tidak terjadi kebuntuan.

3.3.4. Evaluasi Model

Setelah kedua model diterapkan pada dataset, evaluasi dilakukan dengan menggunakan metrik-metrik berikut:

- Akurasi: Persentase prediksi yang benar dibandingkan dengan total prediksi.
- Precision dan Recall: Untuk mengevaluasi kemampuan model dalam mengklasifikasikan mahasiswa yang puas atau Tp.

- F1-Score: Gabungan antara precision dan recall untuk memberikan gambaran keseluruhan performa model.
- Confusion Matrix: Untuk melihat distribusi prediksi dan mengidentifikasi kesalahan klasifikasi yang terjadi.

3.3.5. Perbandingan Hasil

Setelah evaluasi, perbandingan antara Naïve Bayes dan KNN akan dilakukan berdasarkan hasil evaluasi yang diperoleh. Beberapa hal yang akan dibandingkan meliputi:

- Akurasi: Apakah salah satu model lebih akurat dalam memprediksi kepuasan mahasiswa.
- Kecepatan: Waktu yang dibutuhkan untuk melatih dan menguji model.
- Kemampuan Menangani Data Besar: Seberapa baik kedua model dapat menangani dataset yang lebih besar atau lebih kecil.
- Kepraktisan: Kemudahan penerapan model dalam konteks dunia nyata di Universitas Nasional.

4. HASIL DAN PEMBAHASAN

4.1. Deskriptif Dataset

Tahap awal yang dilakukan adalah mengumpulkan data kepuasan mahasiswa terhadap layanan Wi-Fi di Universitas Nasional. Pengumpulan data didapatkan dari survei Google Form yang dibagikan ke media sosial.

4.2. Perbandingan Kinerja Model Klasifikasi

4.2.1. Naïve Bayes

Berikut adalah Confusion matrix berbentuk tabel matriks 5x5 dengan kinerja model klasifikasi pada serangkaian data testing yang nilainya diketahui yang dijelaskan:

Tabel 3. Confusion Matrix Naïve Bayes

	Actual					
Predicted		1	2	3	4	5
	1	20	0	0	0	0
	2	0	23	0	0	0
	3	0	0	13	0	0
	4	0	0	0	22	0
	5	0	0	0	0	22

Confusion matrix di atas menunjukkan hasil klasifikasi model untuk lima kelas yang berbeda, dengan sumbu vertikal sebagai label aktual dan sumbu horizontal sebagai prediksi model. Nilai diagonal utama (20, 23, 13, 22, 22) menunjukkan jumlah data yang diklasifikasikan dengan benar untuk masing-masing kelas, sementara nilai di luar diagonal mencerminkan kesalahan prediksi. Berikut perhitungan manual dari Accuracy, Precision, Recall, dan F1- Score dari klasifikasi menggunakan Naïve Bayes:

a. Accuracy

$$\text{Accuracy} = \frac{\sum TP}{\sum \text{Total Data}} = \frac{20 + 23 + 13 + 22 + 22}{20 + 23 + 13 + 22 + 22} = 1.0$$

b. Precision

$$\text{Precision}_i = \frac{TP_i}{TP_i + \sum FP_i}$$

- Precision untuk kelas 1:

$$= \frac{20}{20 + 0 + 0 + 0 + 0} = \frac{20}{20} = 1.0$$
- Precision untuk kelas 2:

$$= \frac{23}{0 + 23 + 0 + 0 + 0} = \frac{23}{23} = 1.0$$
- Precision untuk kelas 3:

$$= \frac{13}{0 + 0 + 13 + 0 + 0} = \frac{13}{13} = 1.0$$
- Precision untuk kelas 4:

$$= \frac{22}{0 + 0 + 0 + 22 + 0} = \frac{22}{22} = 1.0$$
- Precision untuk kelas 5:

$$= \frac{22}{0 + 0 + 0 + 0 + 22} = \frac{22}{22} = 1.0$$

c. Recall

$$\text{Recall}_i = \frac{TP_i}{TP_i + \sum FN_i}$$

- Recall untuk kelas 1:

$$= \frac{20}{20 + 0 + 0 + 0 + 0} = \frac{20}{20} = 1.0$$
- Recall untuk kelas 2:

$$= \frac{23}{0 + 23 + 0 + 0 + 0} = \frac{23}{23} = 1.0$$
- Recall untuk kelas 3:

$$= \frac{13}{0 + 0 + 13 + 0 + 0} = \frac{13}{13} = 1.0$$
- Recall untuk kelas 4:

$$= \frac{22}{0 + 0 + 0 + 22 + 0} = \frac{22}{22} = 1.0$$
- Recall untuk kelas 5:

$$= \frac{22}{0 + 0 + 0 + 0 + 22} = \frac{22}{22} = 1.0$$

d. F1-Score

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

- F1-Score untuk kelas 1:

$$= 2 \times \frac{1.0 \times 1.0}{1.0 + 1.0} = 1.00$$
- F1-Score untuk kelas 2:

$$= 2 \times \frac{1.0 \times 1.0}{1.0 + 1.0} = 1.00$$
- F1-Score untuk kelas 3:

$$= 2 \times \frac{1.0 \times 1.0}{1.0 + 1.0} = 1.00$$
- F1-Score untuk kelas 4:

$$= 2 \times \frac{1.0 \times 1.0}{1.0 + 1.0} = 1.00$$
- F1-Score untuk kelas 5:

$$= 2 \times \frac{1.0 \times 1.0}{1.0 + 1.0} = 1.00$$

4.2.2. K-Nearest Neighbors (K=3)

Berikut adalah Confusion matrix berbentuk tabel matriks 5x5 dengan kinerja metode K-Nearest Neighbors (K=3) yang dijelaskan:

Tabel 4. Confusion Matrix K-3

	Actual					
		1	2	3	4	5
Predicted	1	20	0	0	0	0
	2	0	18	1	3	1
	3	0	1	12	0	0
	4	0	3	0	19	0
	5	0	2	2	0	18

Confusion matrix di atas menunjukkan hasil klasifikasi model untuk lima kelas yang berbeda, dengan sumbu vertikal sebagai label aktual dan sumbu horizontal sebagai prediksi model. Nilai diagonal utama (20, 18, 12, 19, 18) menunjukkan jumlah data yang diklasifikasikan dengan benar untuk masing-masing kelas, sementara nilai di luar diagonal mencerminkan kesalahan prediksi.

4.2.3. K-Nearest Neighbors (K=5)

Berikut adalah Confusion matrix berbentuk tabel matriks 5x5 dengan kinerja K-Nearest Neighbors (K=5) yang dijelaskan:

Tabel 5. Confusion Matrix K-5

	Actual					
		1	2	3	4	5
Predicted	1	16	4	0	0	0
	2	1	16	1	4	1
	3	0	1	12	0	0
	4	0	6	0	16	0
	5	1	4	0	0	17

Confusion matrix di atas menunjukkan hasil klasifikasi model untuk lima kelas yang berbeda, dengan sumbu vertikal sebagai label aktual dan sumbu horizontal sebagai prediksi model. Nilai diagonal utama (16, 16, 12, 16, 17) menunjukkan jumlah data yang diklasifikasikan dengan benar untuk masing-masing kelas, sementara nilai di luar diagonal mencerminkan kesalahan prediksi.

4.2.4. K-Nearest Neighbors (K=7)

Berikut adalah Confusion matrix berbentuk tabel matriks 5x5 dengan kinerja metode Naïve Bayes yang dijelaskan:

Tabel 6. Confusion Matrix K-7

	Actual					
		1	2	3	4	5
Predicted	1	17	3	0	0	0
	2	2	20	0	1	0
	3	0	0	13	0	0
	4	0	2	0	20	0
	5	0	6	0	0	16

Confusion matrix di atas menunjukkan hasil klasifikasi model untuk lima kelas yang berbeda, dengan sumbu vertikal sebagai label aktual dan sumbu horizontal sebagai prediksi model. Nilai diagonal utama (17, 20, 13, 20, 16) menunjukkan jumlah data yang diklasifikasikan dengan benar untuk masing-masing kelas, sementara nilai di luar diagonal

mencerminkan kesalahan prediksi. Berikut perhitungan manual dari Accuracy, Precision, Recall, dan F1- Score dari klasifikasi menggunakan KNN :

a. Accuracy

$$\begin{aligned} \text{Accuracy} &= \frac{\sum TP}{\sum \text{Total Data}} \\ &= \frac{17 + 3 + 0 + 0 + 0 + 2 + 20 + 0 + 1 + 0 + 0 + 0 + 13 + 0 + 0 + 0 + 2 + 0 + 20 + 0 + 0 + 6 + 0 + 0 + 16}{17 + 20 + 13 + 20 + 16} \\ &= \frac{86}{100} = 0.86 \end{aligned}$$

b. Precision

$$\text{Precision}_i = \frac{TP_i}{TP_i + \sum FP_i}$$

- Precision untuk kelas 1: $\frac{17}{17 + 2 + 0 + 0 + 0} = \frac{17}{19} = 0.89$
- Precision untuk kelas 2: $\frac{20}{3 + 20 + 0 + 2 + 6} = \frac{20}{31} = 0.65$
- Precision untuk kelas 3: $\frac{13}{0 + 0 + 13 + 0 + 0} = \frac{13}{13} = 1.0$
- Precision untuk kelas 4: $\frac{20}{0 + 1 + 0 + 20 + 0} = \frac{20}{21} = 0.95$
- Precision untuk kelas 5: $\frac{16}{0 + 0 + 0 + 0 + 16} = \frac{16}{16} = 1.0$

c. Recall

$$\text{Recall}_i = \frac{TP_i}{TP_i + \sum FN_i}$$

- Recall untuk kelas 1: $\frac{17}{17 + 3 + 0 + 0 + 0} = \frac{17}{20} = 0.85$
- Recall untuk kelas 2: $\frac{20}{2 + 20 + 0 + 1 + 0} = \frac{20}{23} = 0.87$
- Recall untuk kelas 3: $\frac{13}{0 + 0 + 13 + 0 + 0} = \frac{13}{13} = 1.0$
- Recall untuk kelas 4: $\frac{20}{0 + 2 + 0 + 20 + 0} = \frac{20}{22} = 0.91$
- Recall untuk kelas 5: $\frac{16}{0 + 6 + 0 + 0 + 16} = \frac{16}{22} = 0.73$

d. F1-Score

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

- F1-Score untuk kelas 1: $2 \times \frac{0.89 \times 0.85}{0.89 + 0.85} = 0.88$
- F1-Score untuk kelas 2: $2 \times \frac{0.65 \times 0.87}{0.65 + 0.87} = 0.74$
- F1-Score untuk kelas 3: $2 \times \frac{1.0 \times 1.0}{1.0 + 1.0} = 1.0$
- F1-Score untuk kelas 4: $2 \times \frac{0.95 \times 0.91}{0.95 + 0.91} = 0.93$
- F1-Score untuk kelas 5: $2 \times \frac{1.0 \times 0.73}{1.0 + 0.73} = 0.84$

Dari penjelasan diatas dapat dirangkum sebagai berikut:

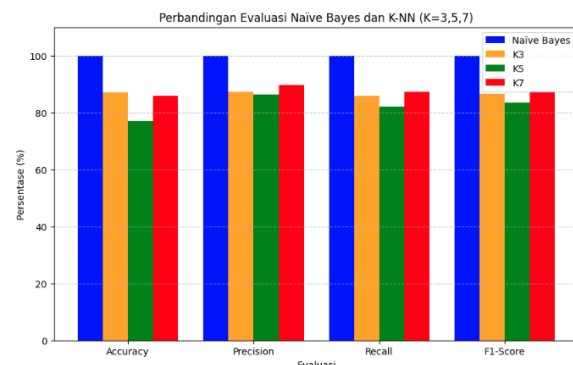
Tabel 6. Tabel perbandingan hasil

Evaluasi	Naïve Bayes	K3	K5	K7
Accuracy	100%	87%	77%	86%
Precision	100%	87.2%	86.4%	89.8%
Recall	100%	86%	82%	87.2%
F1-Score	100%	86.6%	83.6%	87.6%

Hasil evaluasi performa algoritma Naïve Bayes dan K-Nearest Neighbors (K-NN) dengan jumlah tetangga (K) yang berbeda, yaitu K=3, K=5, dan K=7, berdasarkan empat metrik: Accuracy, Precision, Recall, dan F1-Score. Naïve Bayes menunjukkan performa terbaik dengan nilai 100% pada semua metrik, menandakan bahwa model ini mampu mengklasifikasikan data tanpa kesalahan dalam skenario yang diuji. Sementara itu, K-NN dengan K=3 memberikan hasil terbaik dibandingkan dengan K=5 dan K=7, dengan Accuracy 87%, Precision 87.2%, Recall 86%, dan F1-Score 86.6%. Ketika nilai K meningkat, performa model menurun, terlihat dari Accuracy K=5 yang turun menjadi 77% dengan Precision 86.4% dan Recall 82%, sementara K=7 memiliki Accuracy 86%, Precision 89.8%, dan Recall tetap di 87.6%. Secara keseluruhan, hasil ini menunjukkan bahwa Naïve Bayes lebih unggul dalam klasifikasi dibandingkan dengan K-NN, sementara efektivitas K-NN sangat bergantung pada pemilihan nilai K yang optimal.

4.3. Visualisasi Data

Berikut adalah Diagram Perbandingan Evaluasi Naïve Bayes dan K-Nearest Neighbors:



Gambar 2. Perbandingan Diagram Batang

Diagram batang di atas membandingkan hasil evaluasi antara algoritma Naïve Bayes dan K-Nearest Neighbors (K-NN) dengan nilai K=3, K=5, dan K=7 berdasarkan empat metrik: Accuracy, Precision, Recall, dan F1-Score. Hasil menunjukkan bahwa Naïve Bayes memiliki performa tertinggi dengan nilai 100% di semua metrik, sementara K-NN menunjukkan penurunan kinerja seiring perubahan nilai K. Pada K=3, Accuracy mencapai 87%, Precision 87.2%, Recall 86%, dan F1-Score 86.6%, yang masih lebih baik dibandingkan K=5 dan K=7. Pada K=5, performa menurun lebih jauh dengan Accuracy 77%, Precision 86.4%, Recall 82%, dan F1-Score 83.6%, sedangkan K=7 memiliki Accuracy 86%, Precision 89.8%, dan Recall 87.6%.

K=7 sangat membaik dengan Accuracy 86%, Precision 89.8%, Recall 87.2%, dan F1-Score 87.6%. Secara keseluruhan, Naïve Bayes unggul dalam klasifikasi, sedangkan performa K-NN bergantung pada pemilihan nilai K, dengan K=7 memberikan hasil terbaik di antara variasi K.

5. KESIMPULAN DAN SARAN

Penelitian ini menunjukkan bahwa Naïve Bayes memiliki performa lebih unggul dalam klasifikasi kepuasan mahasiswa terhadap layanan Wi-Fi dengan akurasi 100% di semua skenario. Sementara itu, KNN mencapai performa terbaik pada K=7 dengan akurasi 86%, namun cenderung menurun saat jumlah data uji meningkat. Naïve Bayes lebih stabil dan cepat, tetapi rentan terhadap overfitting jika asumsi distribusi data tidak terpenuhi. Sebaliknya, KNN lebih fleksibel dalam menangani pola data tetapi sangat bergantung pada pemilihan nilai K dan jumlah data latih yang cukup besar.

Penelitian selanjutnya dapat mengembangkan metode hybrid Naïve Bayes dan KNN untuk meningkatkan akurasi, serta menggunakan dataset yang lebih besar dan beragam. Optimalisasi parameter K pada KNN, penerapan seleksi fitur, dan analisis faktor eksternal seperti jam sibuk atau lokasi akses Wi-Fi juga dapat dieksplorasi untuk hasil yang lebih akurat.

DAFTAR PUSTAKA

- [1] T. Wahyuningsih, D. Manongga, I. Sembiring, and S. Wijono, "Comparison of Effectiveness of Logistic Regression, Naive Bayes, and Random Forest Algorithms in Predicting Student Arguments," in *Procedia Computer Science*, Elsevier B.V., 2024, pp. 349–356. doi: 10.1016/j.procs.2024.03.014.
- [2] Z. Turgut, S. Üstebay, G. Zeynep Gürkaş Aydın, and A. Sertbaş, "Deep learning in indoor localization using WiFi," in *Lecture Notes in Electrical Engineering*, Springer Verlag, 2019, pp. 101–110. doi: 10.1007/978-981-13-0408-8_9.
- [3] F. Yu *et al.*, "5 G WiFi Signal-Based Indoor Localization System Using Cluster k -Nearest Neighbor Algorithm," *Int J Distrib Sens Netw*, 2019, doi: 10.1155/2014/247525.
- [4] Y. T. Samuel and K. Dewi, "Penggunaan Metode NAÏVE BAYES Dalam Mengukur Tingkat Kepuasan Pengguna Terhadap Online System Universitas Advent Indonesia The Use of Naïve Bayes Method in Measuring User's Satisfaction With Adventist University of Indonesia's Online System," Oct. 2019. [Online]. Available: <https://www.online.unai.edu>.
- [5] Z. Q. Xingbin Ge, 2016 *7th IEEE International Conference on Software Engineering and Service Science (ICSESS)*. IEEE, 2019.
- [6] M. Siddik, R. Noratama Putri, and Y. Desnelita, "CLASSIFICATION OF STUDENT SATISFACTION ON HIGHER EDUCATION SERVICES USING NAÏVE BAYES ALGORITHM," *Journal of Information Technology and Computer Science (INTECOMS)*, vol. 3, no. 2, 2020.
- [7] M. K. M. F. A. S. K. Khurrotul Aeni, "603-Article Text-1574-1-10-20201218," Dec. 2020.
- [8] Q. Kiswara, M. Safii, S. R. Andani, M. R. Lubis, and R. Renaldi, "Implementasi Algoritma C4.5 Untuk Mengukur Tingkat Kepuasan Mahasiswa yang Berlangganan Wifi Indihome," *TIN: Terapan Informatika Nusantara*, vol. 4, no. 9, pp. 620–627, Feb. 2024, doi: 10.47065/tin.v4i9.4918.
- [9] J. M. Ima Wardani, "PENGARUH+HARGA,+FREE+WIFI+DAN+FASILITAS+TERHADAP+KEPUASAN+PELANGGAN+PADA+KEDAI+COFFEE+JMP+PAHLAWAN+LAMONGAN," 2020.
- [10] P. A. M. Zidane, R. W. P. P. Zer, and I. Gunawan, "JURNAL MEDIA INFORMATIKA [JUMIN] Penerapan Data Mining Naïve Bayes Dalam Klasifikasi Kepuasan Mahasiswa Berlangganan WiFi Indihome," 2022.
- [11] W. Lan and H. Li, "A Modified KNN Indoor WiFi Localization Method with K-median Cluster," in *IOP Conference Series: Materials Science and Engineering*, Institute of Physics Publishing, Aug. 2019. doi: 10.1088/1757-899X/608/1/012008.
- [12] Z. E. Rasjid and R. Setiawan, "Performance Comparison and Optimization of Text Document Classification using k-NN and Naïve Bayes Classification Techniques," in *Procedia Computer Science*, Elsevier B.V., 2019, pp. 107–112. doi: 10.1016/j.procs.2017.10.017.
- [13] D. Olivita, Y. Vitriani, J. Teknik Informatika, F. Sains dan Teknologi, U. H. Sultan Syarif Kasim Riau Jl Soebrantas No, and S. Baru, "Perbandingan Klasifikasi Tugas Akhir Mahasiswa Jurusan Teknik Informatika Menggunakan Metode Naïve Bayes Classifier dan K-Nearest Neighbor," *Jurnal Sains, Teknologi dan Industri*, vol. 14, no. 1, pp. 79–85, 2019, [Online]. Available: <http://ejournal.uin-suska.ac.id/index.php/sitekin>