

Perbandingan Algoritma *K-Nearest Neighbor* dan *Naive Bayes* untuk Klasifikasi FoMO Pengguna Media Sosial

DOI: <http://dx.doi.org/10.35889/progresif.v21i2.2784>

Creative Commons License4.0 (CC BY –NC)



Muhammad Haromaen^{1*}, Marthin Piskana², M. Bucci Ryando³, Wira Hadinata⁴

Teknik Informatika, Institut Teknologi dan Bisnis Bina Sarana Global, Tangerang, Indonesia

*e-mail *Corresponding Author*: haromaen23@gmail.com

Abstract

The intensive use of social media among students poses a risk of triggering Fear of Missing Out (FoMO), which negatively affects mental health and learning focus. This study aims to develop a classification model to detect FoMO tendencies among students at SMAN 11 Kabupaten Tangerang. A quantitative approach was used, employing the K-Nearest Neighbor (KNN) and Naive Bayes algorithms. The analyzed variables include gender, duration of social media use, access frequency, desire to stay updated, and its impact on productivity. Data were collected from 244 respondents and processed through pre-processing, modeling, and evaluation stages. Validation results show that KNN achieved the highest accuracy at 94.69%, while Naive Bayes reached 93.06%. These findings indicate that KNN is more effective in detecting FoMO tendencies based on numerical data and has the potential to support early intervention efforts in educational settings.

Keywords: *Fear of Missing Out; K-Nearest Neighbor; Social Media; Classification; Naive Bayes*

Abstrak

Penggunaan media sosial secara intensif di kalangan pelajar berisiko memunculkan gejala *Fear of Missing Out* (FoMO), yang berdampak negatif pada kesehatan mental dan fokus belajar. Penelitian ini bertujuan untuk mengembangkan model klasifikasi kecenderungan FoMO pada pelajar SMAN 11 Kabupaten Tangerang. Metode yang digunakan adalah pendekatan kuantitatif dengan algoritma *K-Nearest Neighbor* (KNN) dan *Naive Bayes*. Variabel yang dianalisis meliputi jenis kelamin, durasi penggunaan media sosial, frekuensi akses, keinginan untuk tetap update, dan pengaruh terhadap produktivitas. Data dikumpulkan dari 244 responden dan diproses melalui pre-processing, modeling, dan evaluasi. Hasil validasi menunjukkan bahwa KNN menghasilkan akurasi tertinggi sebesar 94,69%, sementara *Naive Bayes* mencapai 93,06%. Temuan ini menunjukkan bahwa KNN lebih efektif untuk mendeteksi kecenderungan FoMO berbasis data numerik dan berpotensi mendukung pengembangan intervensi dini dalam konteks pendidikan.

Kata kunci: *Fear of Missing Out; K-Nearest Neighbor; Media Sosial; Klasifikasi; Naive Bayes*

1. Pendahuluan

Perkembangan teknologi digital di Indonesia menunjukkan tren peningkatan yang signifikan setiap tahunnya, terutama dalam hal jumlah pengguna internet dan media sosial. Hingga Januari 2025, jumlah pengguna internet di Indonesia telah mencapai 221 juta jiwa atau setara dengan 79,5% dari total populasi, dengan sekitar 143 juta di antaranya merupakan pengguna aktif media sosial, yakni sekitar 50,2% dari populasi nasional [1], [2]. Peningkatan tersebut menunjukkan bahwa media sosial telah menjadi bagian penting dalam kehidupan sehari-hari masyarakat, khususnya sebagai sarana untuk berkomunikasi, berinteraksi, dan berbagi informasi secara daring. Namun, tingginya intensitas penggunaan media sosial juga menimbulkan sejumlah dampak negatif, salah satunya adalah fenomena *Fear of Missing Out* (FoMO), yaitu perasaan cemas karena takut tertinggal dari informasi atau aktivitas sosial yang dilakukan orang lain [3], [4]. Gejala ini semakin relevan untuk diteliti mengingat dampaknya yang signifikan terhadap kesehatan mental dan perilaku sosial, khususnya pada remaja [5].

Salah satu kelompok yang rentan mengalami FoMO adalah pelajar SMA, yang berada pada fase perkembangan psikososial yang dinamis. Berdasarkan observasi awal di SMAN 11 Kabupaten Tangerang, banyak siswa menghabiskan waktu berjam-jam setiap hari di media sosial, yang mengakibatkan penurunan fokus belajar dan kecemasan sosial. Masalah ini dapat diukur melalui indikator seperti frekuensi penggunaan media sosial, gangguan konsentrasi saat belajar, dan tingkat kecemasan jika tidak mengakses media sosial.

Untuk mendeteksi kecenderungan FoMO secara lebih akurat dan cepat, pendekatan berbasis machine learning menjadi solusi yang menjanjikan. Dua algoritma yang populer digunakan dalam klasifikasi perilaku digital adalah *K-Nearest Neighbor* (KNN) dan *Naïve Bayes*. KNN unggul dalam kemudahan implementasi dan efektif untuk data non-linear [6], sedangkan *Naïve Bayes* lebih efisien dalam pengolahan data besar karena pendekatan probabilitiknya [7]. Sejumlah studi menunjukkan bahwa performa kedua algoritma ini bervariasi tergantung konteks dan karakteristik data [8], [9], [10], [11], sehingga membandingkan keduanya dalam konteks pendeteksian FoMO akan memberikan gambaran metode mana yang lebih sesuai digunakan.

Penelitian ini bertujuan untuk mengembangkan model klasifikasi untuk mendeteksi kecenderungan FoMO pada pelajar SMAN 11 Kabupaten Tangerang menggunakan algoritma KNN dan *Naïve Bayes*. Perbandingan akurasi dan efisiensi dari kedua metode ini akan dilakukan untuk menentukan pendekatan yang optimal. Penelitian ini memberikan dasar empiris bagi guru, psikolog, serta pembuat kebijakan dalam menyusun strategi intervensi yang bertujuan mengurangi pengaruh buruk media sosial terhadap kesehatan mental remaja.

2. Tinjauan Pustaka

Penelitian-penelitian sebelumnya telah banyak mengkaji penerapan algoritma *machine learning*, khususnya *K-Nearest Neighbor* (KNN) dan *Naïve Bayes*, dalam mengklasifikasikan perilaku pengguna terhadap teknologi informasi, termasuk media sosial. [8] melakukan penelitian untuk mengklasifikasikan ulasan pengguna aplikasi *Gapura UB* menggunakan kedua algoritma tersebut. Metode yang digunakan adalah pendekatan *supervised learning* dengan membandingkan performa algoritma terhadap data ulasan yang telah dipreproses. Variabel yang dianalisis meliputi sentimen pengguna dan akurasi klasifikasi. Hasilnya menunjukkan bahwa algoritma *Naïve Bayes* mencatat akurasi sebesar 88,5%, lebih tinggi dibandingkan KNN yang mencapai 84,8%.

Penelitian oleh [9] mengangkat analisis sentimen terhadap aplikasi pembelajaran *Zenius* menggunakan algoritma KNN dan *Naïve Bayes*. Penelitian ini memanfaatkan metode text mining dengan data berupa komentar pengguna. Variabel utama yang dianalisis adalah label sentimen dan akurasi klasifikasi. Hasil menunjukkan bahwa algoritma KNN unggul secara signifikan dengan akurasi 100%, sedangkan *Naïve Bayes* hanya mencapai 88,41%. Ini menunjukkan bahwa performa algoritma sangat kontekstual tergantung pada karakteristik data.

Sementara itu, [10] menerapkan kedua algoritma untuk menganalisis sentimen pengguna terhadap aplikasi *Pedulilindungi*. Penelitian ini menggunakan data survei pengguna dan proses klasifikasi dilakukan untuk menentukan tingkat kepuasan terhadap aplikasi. KNN mencatat akurasi 73,33%, lebih tinggi dari *Naïve Bayes* yang hanya mencapai 70,48%. Penelitian ini menekankan pada pentingnya pemilihan algoritma berdasarkan kompleksitas dan jenis data yang digunakan.

Penelitian lain oleh [12] menggunakan KNN dan *Naïve Bayes* untuk klasifikasi perilaku pengguna e-wallet. Metode yang digunakan adalah pemrosesan data transaksi dan survei perilaku konsumen. Dalam penelitian ini, variabel yang digunakan mencakup preferensi pengguna, intensitas penggunaan, serta kepuasan pengguna. Hasil menunjukkan bahwa KNN lebih unggul dengan akurasi rata-rata 89,44%, dibandingkan *Naïve Bayes* yang mencatat 73,03%.

Sebaliknya, penelitian oleh [11] menunjukkan hasil berbeda. Dalam konteks klasifikasi respons guru terhadap pelaksanaan PPPK (Pegawai Pemerintah dengan Perjanjian Kerja), *Naïve Bayes* sedikit lebih unggul dengan akurasi 75,53%, sedangkan KNN hanya mencapai 73,41%. Data yang dianalisis berupa hasil survei dan opini terbuka, dan proses klasifikasi bertujuan untuk mengelompokkan respon guru berdasarkan persepsi mereka.

State of the Art dari penelitian ini terletak pada objek kajian yang lebih spesifik dan pendekatan model yang difokuskan pada deteksi gejala *Fear of Missing Out* (FoMO) pada pelajar pengguna media sosial. Tidak seperti penelitian sebelumnya yang berfokus pada klasifikasi sentimen umum terhadap aplikasi digital, penelitian ini mengembangkan model klasifikasi untuk

mendeteksi kecenderungan psikologis spesifik (FoMO), dengan melibatkan variabel-variabel seperti intensitas penggunaan media sosial, kecemasan sosial, dan gangguan konsentrasi belajar. Selain itu, penelitian ini melakukan perbandingan performa algoritma KNN dan *Naïve Bayes* dalam konteks klasifikasi FoMO secara kuantitatif dan kontekstual. Hal ini menghadirkan *novelty* dalam konteks penggunaan machine learning untuk mendeteksi isu psikologis digital yang relevan dengan kesehatan mental remaja, yang belum banyak diangkat dalam penelitian-penelitian sebelumnya.

3. Metodologi

Penelitian ini menggunakan metode kuantitatif dengan pendekatan eksperimen komparatif, yang bertujuan untuk membandingkan performa algoritma *K-Nearest Neighbor (KNN)* dan *Naïve Bayes* dalam mengklasifikasikan tingkat *FoMO (Fear of Missing Out)* berdasarkan data hasil kuesioner. Berikut *flowchart* struktur metode penelitian:



Gambar 1. Struktur Metode Penelitian

3.1. Pengumpulan Data

Pada tahapan pengumpulan data, peneliti menggunakan kuesioner dalam Google Form yang disebarakan menggunakan *whatsapp* kepada Siswa/I SMAN 11 Kab.Tangerang, data yang terkumpul akan di olah menggunakan perangkat lunak Ms.Excel [13]. Metode ini efektif dalam mendapatkan data primer melalui observasi dan kuesioner dalam penelitian pendidikan dan sosial. Berikut adalah data table 1 deskripsi setiap atribut yang digunakan,

Tabel 1. Atribut

Nama Atribut	Tipe Data	Deskripsi
Jenis Kelamin	Kategori	Jenis kelamin responden, digunakan untuk melihat apakah terdapat perbedaan kecenderungan FoMO antara laki-laki dan perempuan.
Jam Sosmed	Numerik	Rata-rata jumlah jam penggunaan media sosial per hari. Semakin tinggi nilainya, bisa mengindikasikan ketergantungan terhadap media sosial.
Frekuensi Buka	Numerik	Berapa kali dalam sehari responden membuka media sosial. Frekuensi tinggi bisa menjadi indikator adanya kecemasan akan ketinggalan informasi.
Keinginan Update	Numerik (1–10)	Skor keinginan responden untuk tetap mengikuti perkembangan di media sosial. Semakin tinggi skornya, semakin kuat dorongan untuk selalu 'update'.
Produktivitas	Numerik (1–10)	Seberapa besar pengaruh media sosial terhadap tingkat produktivitas responden. Nilai tinggi bisa mengindikasikan dampak negatif terhadap efektivitas kerja/belajar.
Target Klasifikasi (FoMO)	Label Klasifikasi	Label yang akan digunakan untuk klasifikasi pengguna apakah termasuk mengalami FoMO atau tidak (Ya/Tidak).

3.2. Pre-Processing Data

Pada tahap ini, data yang dikumpulkan dari responden melalui kuisisioner yang dibagikan akan di olah untuk mengatasi masalah nilai kosong, duplikasi, atau perbedaan skala. Teknik yang digunakan mencakup imputasi mean atau median untuk data yang hilang, serta normalisasi fitur [14]. Langkah–langkah utama meliputi normalisasi data untuk menyetarakan skala atribut menggunakan rumus:

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)} \tag{1}$$

Dimana x adalah nilai asli, $\min(x)$ adalah nilai minimum dan $\max(x)$ adalah nilai maksimum pada atribut tersebut. Selain itu, dilakukan seleksi fitur dengan menghapus atribut yang tidak relevan untuk memastikan kualitas data yang optimal [15].

3.3. Modeling

Tahapan modeling dilakukan dengan membangun model klasifikasi menggunakan algoritma *K-Nearest Neighbor* (KNN) dan *Naïve Bayes* untuk membandingkan bagaimana kinerja kedua algoritma tersebut.

1) *K-Nearest Neighbor*

KNN adalah algoritma klasifikasi yang menentukan kelas dari suatu data baru berdasarkan mayoritas kelas dari 'k' tetangga terdekatnya [16].

Rumus Jarak *Euclidean*:

Untuk dua titik data $x = (x_1, x_2, \dots, x_n)$ dan $y = (y_1, y_2, \dots, y_n)$ jarak *Euclidean* dihitung sebagai:

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

Keterangan:

- a. $d(x, y)$: jarak
- b. x_i : nilai fitur ke-i dari data latih
- c. y_i : nilai fitur ke-i dari data uji
- d. n : jumlah Fitur
- e.

Langkah ini penting untuk kedekatan terhadap data uji dan data latih [10]. Setelah itu diambil K data terdekat berdasarkan jarak terkecil yang sudah dihitung [17]. Kemudian dilakukan voting mayoritas terhadap kelas dari K tetangga terdekat. Kelas yang paling banyak muncul dijadikan sebagai hasil klasifikasi [5].

2) *Naïve Baiyes*

Naïve Bayes merupakan algoritma klasifikasi yang didasarkan pada perhitungan probabilitas menggunakan Teorema *Bayes*, dengan asumsi bahwa setiap fitur saling bebas atau independen. [18].

Rumus Teorama *Baiyes*:

$$P(C|X) = \frac{P(X|C) \cdot P(C)}{P(X)} \tag{3}$$

Dimana:

- a. $P(C|X)$ adalah probabilitas C diberikan fitur X .
- b. $P(X|C)$ adalah probabilitas X diberikan fitur C .
- c. $P(C)$ adalah probabilitas awal dari kelas C .
- d. $P(X)$ adalah probabilitas awal dari kelas X . [18].

3.4. Evaluasi Model

Evaluasi model bertujuan untuk mengukur kinerja algoritma klasifikasi menggunakan *confusion matrix* untuk mengukur kinerja algoritma klasifikasi yang digunakan, yaitu *K-Nearest Neighbor* (KNN) dan *Naïve Bayes* [19], [20], rumus:

- 1) Akurasi

$$Akurasi = \frac{TP + TN}{TP + TN + FP + FN} \quad (4)$$

- 2) Presisi

$$Presisi = \frac{TP}{TP + FP} \quad (5)$$

- 3) Recall

$$Recall = \frac{TP}{TP + FN} \quad (6)$$

- 4) F1-Score

$$F1 = 2 \cdot \frac{Presisi \cdot Recall}{Presisi + Recall} \quad (7)$$

Hasil ini membantu menilai keefektifan algoritma KNN, sebagaimana dijelaskan dalam penelitian [15] dan [21].

4. Hasil dan Pembahasan

4.1. Pengumpulan Data

Peneliti melakukan pengumpulan data melalui *Google Form* yang telah disebar, di mana data tersebut dikumpulkan dari siswa/i dan mahasiswa/i. Penelitian ini dilakukan di SMAN 11 Kabupaten Tangerang, dengan jumlah data yang berhasil terkumpul sebanyak 244 responden. Data yang diperoleh mencakup beberapa atribut yang telah ditentukan sebelumnya. Metode ini dinilai efektif dalam memperoleh data primer melalui observasi dan kuesioner, terutama dalam konteks penelitian pendidikan dan sosial. Proses analisis, pengolahan data, dan pengembangan model dilakukan menggunakan *Google Colab* dengan memanfaatkan bahasa pemrograman Python. Berikut dataset yang digunakan sesuai dengan atribut yang sudah ditentukan:

Tabel 2. Info Dataset

No	Jenis Kelamin	Jam Sosmed	Frekuensi Buka	Keinginan Update	Produktivitas
1	Perempuan	9	8	1	2
2	Perempuan	5	3	1	5
3	Perempuan	10	3	2	5
4	Perempuan	11	30	8	6
5	Perempuan	3	8	5	5
...
241	Perempuan	6	6	5	5
242	Perempuan	1	18	4	2
243	Laki - laki	3	4	3	4
244	Laki - laki	3	3	1	7
245	Laki - laki	3	4	3	8

Setelah dataset didapatkan selanjutnya dilakukan pelabelan fomo dan tidak fomo menggunakan.

4.2. Pre-processing Data

	Jam Sosmed	Frekuensi Buka	Keinginan Update	Produktivitas
count	244.000000	244.000000	244.000000	244.000000
mean	5.577869	11.069672	4.155738	5.307377
std	2.918312	8.056849	2.248074	2.225874
min	1.000000	1.000000	1.000000	1.000000
25%	3.000000	4.000000	3.000000	4.000000
50%	5.000000	8.000000	4.000000	5.000000
75%	7.000000	17.000000	5.000000	7.000000
max	13.000000	30.000000	10.000000	10.000000

Gambar 2. Cek Missing Value

Di tahap *Pre-Processing* Data ini, penulis melakukan cek awal apakah ada missing value atau data non atau null, proses ini bertujuan untuk mengetahui apakah data-data akan digunakan telah baik kualitasnya dan dapat dipercaya.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 244 entries, 0 to 243
Data columns (total 7 columns):
#   Column                Non-Null Count  Dtype
---  -
0   Nama Lengkap          244 non-null    object
1   Kelas                 244 non-null    object
2   Jenis Kelamin         244 non-null    object
3   Jam Sosmed            244 non-null    int64
4   Frekuensi Buka        244 non-null    int64
5   Keinginan Update      244 non-null    int64
6   Produktivitas         244 non-null    int64
dtypes: int64(4), object(3)
memory usage: 13.5+ KB
```

Gambar 3. Cek Data Null

Dapat dilihat dari gambar tersebut bahwa data digunakan tidak ada missing value dan null data, yaitu artinya kualitas data untuk klasifikasi pengguna media sosial ini dengan kecenderungan fomo ini baik dan dapat diandalkan.

```
[37] df.duplicated(keep=False).sum()
np.int64(0)
```

Gambar 4. Pengecekan Data Duplikat

Selanjutnya penulis melakukan cek duplikasi data dan tidak ada data duplikat pada data hasil kuesioner tersebut sehingga tidak ada data duplikat apa-apa untuk dihapus pada proses ini.

```
X_train, X_test, y_train, y_test = train_test_split(X_scaled, y_encoded, test_size=0.2, random_state=42)
print("Train Size:", len(X_train))
print("Test Size:", len(X_test))
Train Size: 196
Test Size: 49
```

Gambar 5. Split Data

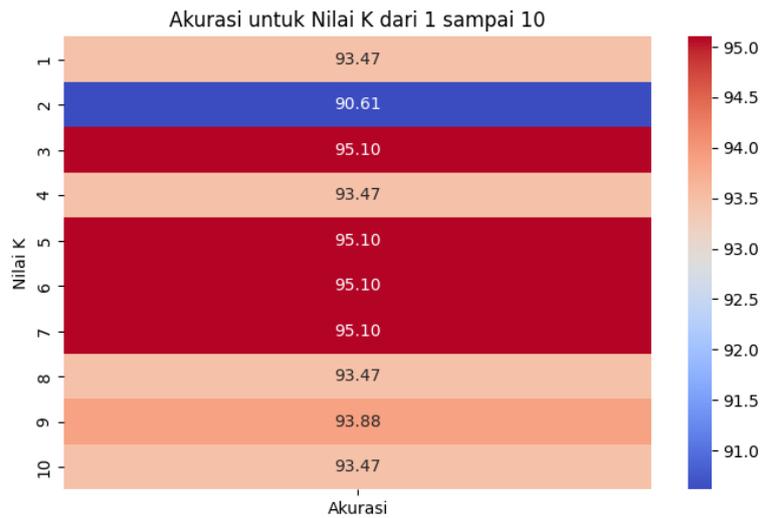
Proses membagi data dilakukan menggunakan fungsi `train_test_split` dari library `scikit-learn`. Data dibagi menjadi data latih 80% dan data uji 20% dengan menggunakan parameter `random_state=42` agar proses pembagian tetap konsisten. Hasilnya, data latih ada sebanyak 194

sampel dan data uji ada sebanyak 49 sampel, sama-sama dengan 10 atribut. Tujuan dari proses ini adalah untuk mempersiapkan data untuk membangun dan menguji.

4.3. Modeling

Pada tahapan modeling, proses eksekusi kedua algoritma klasifikasi, yaitu *Naive Bayes* dan *K-Nearest Neighbor* (KNN), digunakan terhadap data pengguna media sosial yang telah diproses. Di samping itu, ada visualisasi untuk memudahkan pemahaman data dan hasil.

1) Penentuan Parameter



Gambar 6. Heatmap Akurasi

Gambar di atas menampilkan heatmap akurasi dari algoritma *K-Nearest Neighbor* (KNN) berdasarkan variasi nilai *K* dari 1 hingga 10. Dari visualisasi tersebut, terlihat bahwa akurasi tertinggi sebesar 95.10% dicapai pada nilai *K* = 3, 5, 6, dan 7. Sementara itu, nilai *K* = 2 menunjukkan akurasi terendah sebesar 90.61%, menandakan bahwa pemilihan nilai *K* sangat berpengaruh terhadap performa algoritma KNN dalam proses klasifikasi.

Pada tahap ini, algoritma *K-Nearest Neighbor* digunakan dengan mencari nilai *K* terbaik yang memberikan akurasi tertinggi. Sedangkan untuk metode *Naive Bayes*, digunakan pendekatan distribusi probabilitas Gaussian yang memudahkan dalam proses pengklasifikasian. Kedua metode ini diterapkan untuk menentukan klasifikasi pengguna media sosial yang memiliki kecenderungan mengalami *Fear of Missing Out* (FoMO).

Dengan hasil tersebut, pemilihan nilai *K* yang optimal dalam KNN menjadi penting untuk meningkatkan akurasi prediksi, sementara penggunaan distribusi Gaussian pada *Naive Bayes* memberikan pendekatan probabilistik yang sederhana namun efektif untuk klasifikasi.

2) Perhitungan Jarak

```
[ ] distance = np.sqrt(np.sum(np.square(np.array(X_test.shape) - np.array(X_train.shape))))
print("Euclidean Distance:", distance)

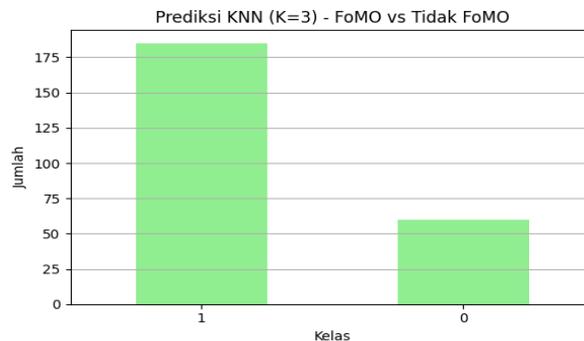
↩ Euclidean Distance: 147.0
```

Gambar 7. Hasil Euclidean

Berdasarkan hasil perhitungan menggunakan Python, diperoleh nilai jarak Euclidean antara bentuk dimensi dari data uji dan data latih sebesar 147.0, seperti ditunjukkan pada Gambar 6. Nilai ini selanjutnya digunakan sebagai dasar untuk menentukan kedekatan data uji dengan data latih dalam proses klasifikasi menggunakan algoritma *K-Nearest Neighbor* (KNN).

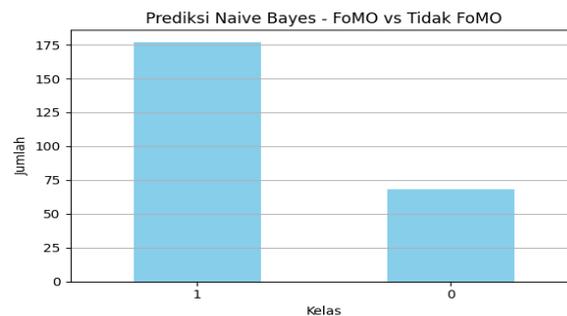
3) Prediksi Kelas

Di tahapan ini, algoritma yang digunakan akan melakukan prediksi terhadap kelas target secara berdasarkan karakteristik entry data yang telah melewati proses training. Berikut adalah cara masing-masing algoritma melakukan prediksi.



Gambar 8. Prediksi Berdasarkan K Terbaik

Pada metode KNN dengan nilai K=3 (nilai terbaik yang diperoleh dari proses validasi), hasil prediksi menunjukkan bahwa sebanyak 185 data (atau sekitar 75,51%) dikategorikan ke dalam kelas FoMO (label 1). Sementara itu, sebanyak 60 data (atau 24,49%) diklasifikasikan sebagai Tidak FoMO (label 0). Visualisasi dari hasil prediksi ini dapat dilihat pada Gambar 8, yang menampilkan dominasi jumlah individu yang masuk dalam kategori FoMO.



Gambar 9. Prediksi Berdasarkan Pendekatan Probabilistik

Sementara itu, dengan pendekatan *Naive Bayes* yang mengandalkan prinsip probabilistik, diperoleh hasil yang sedikit berbeda. Dari total data yang dianalisis, sebanyak 177 data (sekitar 72,24%) diprediksi masuk dalam kategori FoMO, sedangkan 68 data (sekitar 27,76%) diklasifikasikan sebagai Tidak FoMO. Hasil ini divisualisasikan pada Gambar 9, yang juga menunjukkan kecenderungan sebagian besar individu termasuk dalam kategori FoMO.

4.4 Evaluasi Model

	Train Size	Test Size
Fold 1	196	49
Fold 2	196	49
Fold 3	196	49
Fold 4	196	49
Fold 5	196	49

Gambar 10. K-Fold Cross Validation

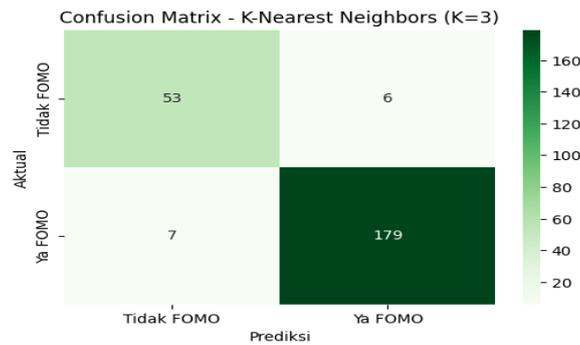
Evaluasi kinerja model dilakukan secara keseluruhan dan fair terhadap semua data yang ada menggunakan strategi *Stratified K-Fold Cross Validation* dengan parameter K = 5. Strategi ini dipilih karena mampu menjaga proporsi target classes (*FoMO* dan Tidak *FoMO*) di masing-masing lipatan (*folds*), sehingga distribusi label di data pelatihan dan data testing tetap seimbang.

Strategi stratifikasi ini sangat berfungsi, terutama dengan menghadapi dataset yang punya distribusi kelas yang tidak seimbang.

Setiap fold ini membagi data menjadi 196 data pelatihan (*training*) dan 49 data pengujian (*testing*). Dengan lima *fold*, keseluruhan data pun berpeluang sama untuk menjadi sebagian dari data uji sebanyak satu kali, dan sebagian dari data pelatihan sebanyak empat kali. Gambar 3 menunjukkan pembagian data di setiap fold.

1) Confusion Matrix

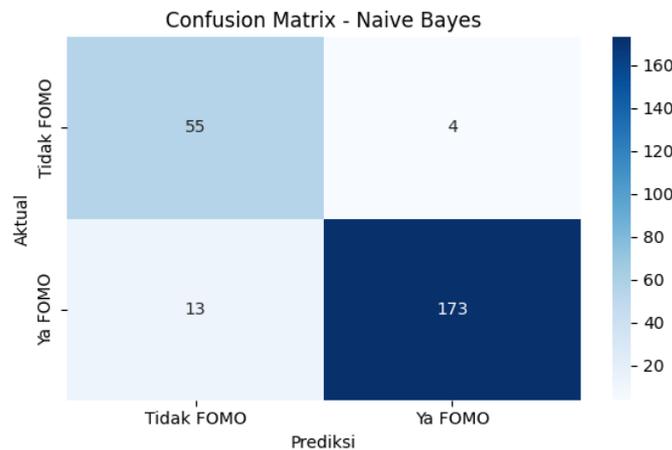
a. KNN



Gambar 11. Confusion Matrix KNN

Berdasarkan gambar di atas, model *K-Nearest Neighbor* (K=3) menunjukkan performa yang sangat baik dalam mengklasifikasikan data FoMO. Dari 59 data aktual yang termasuk kategori Tidak FoMO, sebanyak 53 data berhasil diklasifikasikan dengan benar, sementara 6 data lainnya salah diklasifikasikan sebagai Ya FoMO. Sementara itu, dari 186 data aktual yang tergolong Ya FoMO, sebanyak 179 data berhasil diklasifikasikan secara tepat, dan hanya 7 data yang salah diklasifikasikan sebagai Tidak FoMO. Hasil ini mengindikasikan bahwa KNN (K=3) memiliki tingkat akurasi dan presisi yang tinggi dalam mengenali kedua kategori tersebut.

b. Naïve Baiyes



Gambar 12. Confusion Matrix Naïve Baiyes

Model *Naive Bayes* juga memberikan hasil klasifikasi yang cukup baik, meskipun sedikit berada di bawah performa KNN. Dari 59 data aktual Tidak FoMO, sebanyak 55 data diklasifikasikan secara benar dan 4 data salah diklasifikasikan sebagai Ya FoMO. Sementara untuk kategori Ya FoMO, dari total 186 data, sebanyak 173 data berhasil diklasifikasikan dengan benar, sedangkan 13 data salah diklasifikasikan sebagai Tidak FoMO. Meskipun masih terdapat kesalahan klasifikasi, khususnya pada kategori Ya FoMO, model *Naive Bayes* tetap menunjukkan performa yang cukup andal dalam mendeteksi FoMO.

2) Hasil Klasifikasi dan Perbandingan Algoritma KNN dan *Naïve Baiyes*

Tabel 3. Hasil Klasifikasi KNN dan *Naïve Baiyes*

Metrik	KNN	<i>Naïve Baiyes</i>
Akurasi	94.69%	93.06%
Presisi	94.73%	93.68%
Recall	94.69%	93.06%
F1-Score	94.71%	93.22%

Dari evaluasi performa kedua algoritma dalam mengklasifikasikan data, yaitu *K-Nearest Neighbor* (KNN) dan *Naive Bayes*, dapat disimpulkan bahwa keduanya bekerja dengan sangat baik, meskipun terdapat perbedaan yang mencolok dalam hal akurasi dan konsistensi.

Algoritma KNN memiliki tingkat akurasi 94,69%, sedikit lebih tinggi dibandingkan 93,06% yang diperoleh oleh *Naive Bayes*. Ini menunjukkan bahwa KNN adalah algoritma yang lebih efisien untuk klasifikasi data secara keseluruhan. Dari segi presisi, KNN juga unggul dengan skor 94,73% dibandingkan *Naive Bayes* yang memperoleh 93,68%, artinya prediksi positif yang dibuat oleh KNN lebih akurat dan memiliki tingkat kesalahan klasifikasi yang lebih rendah.

Dalam hal recall, KNN juga unggul dengan nilai 94,69% dibandingkan 93,06% milik *Naive Bayes*. Ini menunjukkan bahwa KNN lebih efektif dalam menangkap seluruh data yang seharusnya diklasifikasikan sebagai positif, sehingga mengurangi kemungkinan informasi penting yang terlewatkan. *F1-Score* sebagai indikator keseimbangan antara presisi dan recall juga mencerminkan keunggulan KNN dengan skor 94,71% dibandingkan 93,22% milik *Naive Bayes*. Hal ini menunjukkan bahwa KNN tidak hanya akurat, tetapi juga konsisten dalam mengklasifikasikan data positif.

Secara umum, meskipun algoritma *Naive Bayes* tetap efisien dan cukup akurat, algoritma KNN terbukti lebih unggul dalam semua metrik evaluasi utama (akurasi, presisi, recall, dan *F1-Score*). Oleh karena itu, KNN merupakan algoritma yang lebih tepat untuk diterapkan dalam konteks dataset dan permasalahan klasifikasi ini.

4.5. Pembahasan

Hasil dari penelitian ini memperlihatkan bahwa algoritma KNN memiliki performa lebih tinggi daripada *Naive Bayes* dalam mengklasifikasikan kecenderungan FoMO dengan akurasi sebesar 94,69% berbanding 93,06%. Temuan ini sejalan dengan hasil penelitian oleh [9] yang juga mencatat bahwa KNN lebih unggul dalam klasifikasi sentimen aplikasi Zenius (100% vs 88,41%). Demikian pula, [10] dan [12] melaporkan performa KNN yang konsisten lebih tinggi dalam konteks klasifikasi pengguna digital.

Di sisi lain, hasil ini sedikit berbeda dari [8] yang menemukan bahwa *Naive Bayes* lebih unggul untuk klasifikasi ulasan aplikasi. Perbedaan ini bisa disebabkan oleh jenis data dan kompleksitas atribut yang digunakan. Dataset dalam penelitian ini bersifat kuantitatif dengan fitur numerik, sehingga lebih cocok dengan pendekatan berbasis jarak seperti KNN. Sementara pada data teks, pendekatan probabilistik *Naive Bayes* bisa lebih unggul.

Dengan demikian, penelitian ini memperkuat temuan sebelumnya tentang efektivitas KNN pada data non-linear dan numerik, serta menambahkan konteks baru karena digunakan untuk klasifikasi kecenderungan psikologis digital (FoMO), bukan hanya sentimen atau preferensi pengguna aplikasi.

5. Simpulan

Berdasarkan pemodelan dan evaluasi hasil klasifikasi, algoritma *K-Nearest Neighbor* (KNN) memperlihatkan performa yang lebih unggul dibandingkan dengan algoritma *Naive Bayes* dalam mendeteksi kecenderungan *Fear of Missing Out* (FoMO) di kalangan pengguna media sosial. KNN memperoleh akurasi yang lebih tinggi (94,69%) dibandingkan *Naive Bayes* (93,06%), serta skor yang lebih baik pada metrik presisi, recall, dan F1-score. Temuan penelitian ini menunjukkan bahwa KNN lebih kuat dan andal dalam mengklasifikasikan pola perilaku psikologis seperti FoMO, berdasarkan atribut data pengguna seperti lama penggunaan media sosial,

keinginan untuk selalu mendapat pembaruan, dan tingkat produktivitas. Oleh karena itu, KNN dapat dianggap sebagai model *machine learning* yang lebih efektif untuk permasalahan klasifikasi serupa yang melibatkan data perilaku dan psikologis.

Dengan kinerja prediktif yang lebih tinggi, algoritma KNN direkomendasikan untuk digunakan dalam studi lanjutan atau aplikasi yang berkaitan dengan profiling psikologis atau klasifikasi perilaku berdasarkan data media sosial. Peneliti dan pengembang di masa depan sebaiknya mempertimbangkan untuk mengintegrasikan model berbasis KNN dalam alat pemantauan kesehatan mental, khususnya untuk mengidentifikasi remaja berisiko yang menunjukkan tanda-tanda FoMO. Selain itu, lembaga pendidikan dan pembuat kebijakan dapat memanfaatkan temuan ini untuk merancang program intervensi dini menggunakan analitik prediktif guna mendukung kesejahteraan siswa dan literasi kesehatan digital.

Daftar Referensi

- [1] S. Kemp, "Digital 2025: Indonesia," Data Reportal. [Online]. Available: <https://datareportal.com/reports/digital-2025-indonesia>
- [2] K. Wisnubroto, "Komitmen Pemerintah Melindungi Anak di Ruang Digital," INDONESIA.GO.ID Portal Informasi Indonesia. [Online]. Available: <https://indonesia.go.id/kategori/editorial/9037/komitmen-pemerintah-melindungi-anak-di-ruang-digital?lang=1>
- [3] B. H. Mulyono, "Pengaruh Fear of Missing Out terhadap Social Connectedness yang Dimediasi oleh Penggunaan Media Sosial," *Bul. Ris. Psikol. dan Kesehat. Ment.*, vol. 1, no. 2, pp. 1190–1198, Aug. 2021, doi: 10.20473/brpkm.v1i2.28660.
- [4] A. Najib and R. P. Rahmawati, "Self-Control: Melindungi Kesehatan Mental Mahasiswa Akibat Fear Of Missing Out (FOMO)," *HASANUDDIN J. Sociol.*, vol. 6, no. 1, pp. 16–24, 2024, [Online]. Available: <http://journal.unhas.ac.id/index.php/HJS/article/view/32433>
- [5] E. Qiudandra, R. Akram, and Novianda, "Sistem Pakar Diagnosa Penyakit Osteoarthritis Dengan Menggunakan Metode K-Nearest Neighbor," *METHOTIKA J. Ilm. Tek. Inform.*, vol. 2, no. 2, pp. 37–48, 2022, [Online]. Available: <http://ojs.fikom-methodist.net/index.php/METHOTIKA>
- [6] A. N. Iffah'da and Anita Desiani, "Implementasi Algoritma K-Nearest Neighbor (K-NN) dan Single Layer Perceptron (SLP) Dalam Prediksi Penyakit Sirosis Biliari Primer," *J. Ilm. Inform.*, vol. 7, no. 1, pp. 65–74, Jul. 2022, doi: 10.35316/jimi.v7i1.65-74.
- [7] S. Sinaga, R. W. Sembiring, and S. Sumarno, "Penerapan Algoritma Naive Bayes untuk Klasifikasi Prediksi Penerimaan Siswa Baru," *J. Mach. ...*, vol. 1, no. 1, pp. 55–64, 2022, [Online]. Available: <https://journal.fkpt.org/index.php/malda/article/view/162%0Ahttps://journal.fkpt.org/index.php/malda/article/download/162/115>
- [8] R. T. S. A. Putri, D. E. Ratnawati, and D. W. Brata, "Perbandingan Naïve Bayes dan K-Nearest Neighbor untuk Analisis Sentimen Aplikasi Gapura UB Berdasarkan Ulasan Pengguna pada Playstore," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 7, no. 1, pp. 229–236, 2023, [Online]. Available: <http://j-ptiik.ub.ac.id>
- [9] T. Abdillah, U. Khaira, and B. F. Hutabarat, "Komparasi Metode Naive Bayes dan K-Nearest Neighbors Terhadap Analisis Sentimen Pengguna Aplikasi Zenius," *J. Process.*, vol. 19, no. 1, pp. 32–44, May 2024, doi: 10.33998/processor.2024.19.1.1596.
- [10] A. P. Wibowo, W. Darmawan, and N. Amalia, "Komparasi Metode Naïve Bayes Dan K-Nearest Neighbor Terhadap Analisis Sentimen Pengguna Aplikasi Pedulilindungi," *IC-Tech*, vol. 17, no. 1, pp. 18–23, Apr. 2022, doi: 10.47775/icttech.v17i1.234.
- [11] F. Fitriani, "Perbandingan Algoritma Naïve Bayes Dan Knn Dalam Analisis Sentimen Masyarakat Terhadap Pelaksanaan PPPK Guru," *Bull. Netw. Eng. Informatics*, vol. 1, no. 1, pp. 18–24, Apr. 2023, doi: 10.59688/bufnets.v1i1.6.
- [12] F. M. Delta Maharani, A. Lia Hananto, S. Shofia Hilabi, F. Nur Apriani, A. Hananto, and B. Huda, "Perbandingan Metode Klasifikasi Sentimen Analisis Penggunaan E-Wallet Menggunakan Algoritma Naïve Bayes dan K-Nearest Neighbor," *METIK J.*, vol. 6, no. 2, pp. 97–103, Dec. 2022, doi: 10.47002/metik.v6i2.372.
- [13] S. Simaremare and R. Siregar, "Effectivity of the Microsoft Excel Application on Student Understanding in Statistics Courses," *Indones. J. Educ. Soc. Humanit.*, vol. 1, no. 2, pp. 9–19, Jun. 2024, doi: 10.62945/ijesh.v1i2.60.
- [14] A. Widiyanti and I. Pratama, "Penanganan Missing Values Dan Prediksi Data Timbunan

- Sampah Berbasis Machine Learning,” *Rabit J. Teknol. dan Sist. Inf. Univrab*, vol. 9, no. 2, pp. 242–251, Jul. 2024, doi: 10.36341/rabit.v9i2.4789.
- [15] S. Prayogo, A. A. Chamid, and A. C. Murti, “Perancangan Sistem Klasifikasi Jenis Bunga Mawar Menggunakan Metode K-Nearest Neighbor (KNN),” *Indones. J. Technol. Informatics Sci.*, vol. 3, no. 2, pp. 52–56, Jun. 2022, doi: 10.24176/ijtis.v3i2.7881.
- [16] F. M. N. Akbar, “Metode KNN (K-Nearest Neighbor) untuk Menentukan Kualitas Air,” *J. Tekno Kompak*, vol. 18, no. 1, pp. 28–40, Feb. 2024, doi: 10.33365/jtk.v18i1.3241.
- [17] N. P. Hapsari and A. Mahmudi, “Penerapan Metode K-Nearest Neighbor Untuk Pengelompokan Ukuran Gamis Customer Berbasis Web,” *J. Inform. Teknol. dan Sains*, vol. 7, no. 1, pp. 367–376, Mar. 2025, doi: 10.51401/jinteks.v7i1.5454.
- [18] R. R. Burhanuddin, “Klasifikasi Penyakit Padi Melalui Citra Daun Menggunakan Metode Naive Bayes,” *J. Inform. dan Tek. Elektro Terap.*, vol. 12, no. 2, pp. 885–892, Apr. 2024, doi: 10.23960/jitet.v12i2.4012.
- [19] K. Anwar, R. Maruf, F. Susanto, and M. B. Ryando, “Analisis Performa Deteksi Penyakit Paru-Paru dengan Model Klasifikasi Gambar Menggunakan LSTM Deep Learning,” *J. Ilm. Univ. Batanghari Jambi*, vol. 25, no. 1, pp. 972–979, Feb. 2025, doi: 10.33087/jjubj.v25i1.5697.
- [20] E. Novianto, A. Hermawan, and D. Avianto, “Klasifikasi Algoritma K-Nearest Neighbor, Naive Bayes, Decision Tree Untuk Prediksi Status Kelulusan Mahasiswa S1,” *Rabit J. Teknol. dan Sist. Inf. Univrab*, vol. 8, no. 2, pp. 146–154, 2023, doi: 10.36341/rabit.v8i2.3434.
- [21] D. I. Muhammad and N. Falih, “Penggunaan K-Nearest Neighbor (KNN) untuk Mengklasifikasi Citra Belimbing Berdasarkan Fitur Warna,” *J. Inform. Ed. ke-17, Nomor 1*, vol. 4221, no. April, pp. 9–16, 2021, doi: <https://doi.org/10.52958/iftk.v17i1.2132>.