

## BAB IV

### HASIL DAN PEMBAHASAN

#### 4.1 Hasil Penelitian

##### 4.1.1 Data Responden

Penelitian ini melibatkan 368 responden yang merupakan pengikut aktif akun *Instagram* resmi Kampus IIB Darmajaya (@darmajayathebest). Responden terdiri dari mahasiswa aktif, calon mahasiswa, serta masyarakat umum yang dipilih berdasarkan kriteria tertentu, yaitu pernah melihat atau berinteraksi dengan konten *Feed*, *Story*, atau *Reels*. Pengumpulan data dilakukan secara online melalui penyebaran kuesioner menggunakan *Google Form*.

	A	B	C	D	E	F	G	H	I
1	Timestamp	Email	Nama	Usia	Status	Gambar pada konten Instagram @darmajayathebest beresolusi tinggi.	Kualitas visual pada setiap ungahan terlihat profesional.	Gambar yang digunakan tidak buram atau pecah.	Warna yang digunakan pada konten terlihat menarik.
2	10/05/2025 2:48:00	itadixru@gmail.com	Daru	19	Ya	5	5	5	
3	10/05/2025 17:32:17	sfebriyanti307@gmail.com	sheila febriyanti	21	Mahasiswa	4	4	4	
4	10/05/2025 17:32:20	aprilreni114@gmail.com	Reni Aprilawati	22	Mahasiswa	3	4	4	
5	10/05/2025 17:33:47	anyagunawan558@gmail.com	Arya Gunawan	23	Mahasiswa	5	5	5	
6	10/05/2025 18:18:53	ntasyaaulina@gmail.com	Naina Hibatasya A	22	Mahasiswa	4	4	4	
7	10/05/2025 18:30:24	ardanfaiz.2111010103@mail.darmajaya.ac.id	ardan	20	Mahasiswa	4	3	3	
8	10/05/2025 18:31:20	wildanbagas491@gmail.com	Wildan	22	Mahasiswa	5	5	5	
9	10/05/2025 18:38:16	zulffroy@gmail.com	zulfi	24	Umum	5	5	5	
10	10/05/2025 19:03:12	alpinka25@gmail.com	Alfingka Mutia Rahma Nitha	22	Mahasiswa	4	4	4	
11	10/05/2025 19:21:09	indahlestanipg@gmail.com	Indah Lestari	21	Mahasiswa	5	5	5	
12	10/05/2025 19:26:50	ranggaaja756@gmail.com	Rangga Syahputra	21	Mahasiswa	3	3	3	
13	10/05/2025 19:27:10	europorom@gmail.com	ziki	23	Mahasiswa	3	3	3	
14	10/05/2025 19:35:34	handikarhama2@gmail.com	Handika Rhma Dwi Ariefky	21	Mahasiswa	4	4	4	
358	12/07/2025 01:33	aradnewartupodia18@gmail.com	ALDO PUTRA WENDARA	21	Mahasiswa	4	4	3	
359	12/07/2025 01:33	ydaluytuebadnma14@mail.darmajaya.ac.id	AMANDA BEAUTY LADY	21	Mahasiswa	2	4	3	
360	12/07/2025 01:33	anggiannaulimafitry56@gmail.com	ANGGI ANNAULIM AFITRY	21	Mahasiswa	3	2	2	
361	12/07/2025 01:33	arironadona95@gmail.com	ARI ROMADONA	20	Mahasiswa	4	4	4	
362	12/07/2025 2:06:57	itmairakzairim78@mail.darmajaya.ac.id	MIRA RIZKI ARIANTI	21	Mahasiswa	3	3	3	
363	12/07/2025 2:07:08	muhannmadabiyuabdullah47@gmail.com	MUHAMMAD ABIYYU ABDULI	20	Mahasiswa	3	2	4	
364	12/07/2025 2:07:19	henietrosno6@gmail.com	HENI RETNO SARI	21	Mahasiswa	4	4	3	
365	12/07/2025 2:07:31	iratsehadin98@gmail.com	INDAH LESTARI	20	Mahasiswa	4	2	3	
366	12/07/2025 2:07:42	anazatin90@gmail.com	ANITA ZANA	21	Mahasiswa	4	2	2	
367	12/07/2025 2:07:55	idnasira51@gmail.com	ARISANDI	20	Mahasiswa	4	2	4	
368	12/07/2025 2:08:07	anoranaluly20@gmail.com	YULIANA RONA	20	Mahasiswa	4	3	2	
369	12/07/2025 2:08:19	rachmatwahyudir4@mail.darmajaya.ac.id	RACHMAT WAHYUDI R	21	Mahasiswa	2	2	3	

Gambar 4. 1 Data Kuesioner

Berdasarkan Gambar 4.1 di atas, dapat dilihat rekapitulasi data mentah (*raw data*) yang berhasil dikumpulkan dari responden melalui *platform Google*

*Form.* Tampilan tersebut memperlihatkan struktur data awal yang terdiri dari stempel waktu (*timestamp*), identitas responden, serta respons jawaban untuk setiap butir pertanyaan kuesioner. Kumpulan data mentah inilah yang selanjutnya akan divalidasi dan diproses pada tahap *pre-processing* untuk memastikan kualitas data sebelum masuk ke tahap pemodelan.

#### 4.1.2 Dataset

Pada penelitian ini, dataset disusun untuk mengukur persepsi serta tingkat *brand awareness* terhadap konten visual yang ditampilkan oleh akun *Instagram* kampus. Setelah seluruh data terkumpul selanjutnya menganalisis sejauh mana konten visual *Instagram* dapat memengaruhi tingkat *brand awareness* terhadap kampus IIB Darmajaya. Berikut dataset yang telah disusun dibawah ini.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 368 entries, 0 to 367
Data columns (total 28 columns):
 #   Column           Non-Null Count  Dtype  
 ---  -- 
 0   Nama            368 non-null    object  
 1   Status           368 non-null    object  
 2   Gambar pada konten Instagram @darmajayathebest beresolusi tinggi. 368 non-null    int64  
 3   Kualitas visual pada setiap unggahan terlihat profesional.        368 non-null    int64  
 4   Gambar yang digunakan tidak buram atau pecah.                   368 non-null    int64  
 5   Warna yang digunakan pada konten terlihat menarik.             368 non-null    int64  
 6   Penggunaan kombinasi warna dalam konten Instagram @darmajayathebest menciptakan tampilan yang menarik secara visual. 368 non-null    int64  
 7   Warna yang digunakan dalam konten Instagram mencerminkan identitas visual kampus (seperti warna logo atau warna khas kampus). 368 non-null    int64  
 8   Setiap postingan memiliki gaya visual yang konsisten.          368 non-null    int64  
 9   Feed Instagram @darmajayathebest terlihat teratur dan selaras secara keseluruhan. 368 non-null    int64  
 10  Desain atau template yang digunakan secara konsisten memudahkan saya mengenali bahwa konten berasal dari instagram @darmajayathebest. 368 non-null    int64  
 11  Informasi dalam gambar atau desain mudah dipahami.            368 non-null    int64  
 12  Tekst dalam gambar dapat dibaca dengan jelas.                  368 non-null    int64  
 13  Konten visual menyampaikan pesan kampus (seperti promosi, informasi kegiatan, atau citra kampus) dengan jelas.       368 non-null    int64  
 14  Konten video Reels yang ditampilkan oleh akun @darmajayathebest menarik untuk ditonton. 368 non-null    int64  
 15  Transisi, musik, dan gaya editing Reels terlihat profesional. 368 non-null    int64  
 16  Konten Reels kampus menyampaikan pesan dengan jelas dan cepat. 368 non-null    int64  
 17  Saya lebih tertarik dengan konten kampus yang berbentuk Reels dari pada gambar biasa atau Feed. 368 non-null    int64  
 18  Saya dapat mengenali konten dari IIB Darmajaya saat melihatnya di Instagram. 368 non-null    int64  
 19  Saya pernah mencari tahu tentang IIB Darmajaya karena melihat konten di Instagram. 368 non-null    int64  
 20  Saya dapat membedakan konten kampus IIB Darmajaya dengan kampus lain. 368 non-null    int64  
 21  Saya merasa melulut konten Instagram, saya jadi lebih mengenal IIB Darmajaya. 368 non-null    int64  
 22  Sejauh mana Anda merasa mengenal dan mengingat IIB Darmajaya sebagai sebuah kampus tanpa bantuan dari iklan atau promosi? 368 non-null    object  
 23  Menurut Anda, apa yang bisa ditingkatkan dari konten Instagram @darmajayathebest ? 368 non-null    object  
 24  Kritik atau saran 344 non-null    object  
 25  Apakah Anda mengikuti akun Instagram resmi IIB Darmajaya (@darmajayathebest)? 368 non-null    object  
 26  Berapa lama Anda sudah mengikuti akun Instagram IIB Darmajaya? 368 non-null    object  
 27  Apakah Anda pernah melakukan interaksi dengan konten Instagram IIB Darmajaya? (boleh pilih lebih dari satu) 368 non-null    object  
dtypes: int64(20), object(8)
memory usage: 80.6+ KB
```

Gambar 4. 2 Dataset yang telah disusun

Berdasarkan Gambar 4.2 di atas, dapat dilihat struktur *dataset* awal yang digunakan dalam penelitian. Gambar tersebut menampilkan informasi bahwa terdapat total 368 data (*entries*) dengan 28 kolom variabel. Tipe data yang terdeteksi didominasi oleh int64 (numerik) untuk data kuesioner skala Likert dan object (teks) untuk data demografi serta opini. Informasi ini memastikan bahwa data telah berhasil dimuat ke dalam sistem sebelum melangkah ke tahap *pre-processing*.

#### 4.1.3 Pre-processing

Dilakukan *pre-processing* untuk memastikan data yang digunakan dalam analisis berkualitas dan bebas dari kesalahan. Tahapan ini sangat penting untuk meningkatkan kualitas hasil analisis serta meminimalkan kesalahan dalam pemrosesan data selanjutnya.

##### 1. Mengganti nama menjadi kategori yang dibutuhkan.

Langkah pertama dalam *pre-processing* adalah melakukan penyesuaian terhadap penamaan kolom atau atribut pada data mentah agar sesuai dengan kebutuhan analisis. Penyesuaian ini mencakup penggantian nama kolom menjadi kategori yang lebih terstruktur, jelas, dan mudah dipahami.

```
# Ganti nama kolom agar lebih mudah dibaca
df = df.rename(columns={
    'Gambar pada konten Instagram @darmajayathebest beresolusi tinggi.': 'kualitas_gambar',
    'Kualitas visual pada setiap ungahan terlihat profesional.': 'kualitas_gambar_2',
    'Gambar yang digunakan tidak buram atau pecah. ': 'kualitas_gambar_3',
    'Transisi, musik, dan gaya editing Reels terlihat profesional. ': 'kualitas_gambar_4',
    'Warna yang digunakan pada konten terlihat menarik. ': 'estetika_warna',
    'Penggunaan Kombinasi warna dalam konten Instagram @darmajayathebest menciptakan tampilan yang menarik secara visual. ': 'estetika_warna_2',
    'Warna yang digunakan dalam konten Instagram mencerminkan identitas visual kampus (seperti warna logo atau warna khas kampus). ': 'estetika_warna_3',
    'Konten video Reels yang ditampilkan oleh akun @darmajayathebest menarik untuk ditonton. ': 'estetika_warna_4',
    'Setiap postingan memiliki gaya visual yang konsisten. ': 'konsistensi_tema',
    'Feed Instagram @darmajayathebest terlihat teratur dan selaras secara keseluruhan. ': 'konsistensi_tema2',
    'Desain atau template yang digunakan secara konsisten memudahkan saya mengenali bahwa konten berasal dari instagram @darmajayathebest. ': 'konsistensi_tema3',
    'Informasi dalam gambar atau desain mudah dipahami. ': 'kejelasan_informasi',
    'Teks dalam gambar dapat dibaca dengan jelas. ': 'kejelasan_informasi2',
    'Konten visual menyampaikan pesan kampus (seperti promosi, informasi kegiatan, atau citra kampus) dengan jelas. ': 'kejelasan_informasi3',
    'Konten Reels kampus menyampaikan pesan dengan jelas dan cepat. ': 'kejelasan_informasi4',
    'Saya dapat mengenali konten dari IIB Darmajaya saat melihatnya di Instagram. ': 'Mengenali konten dari IIB Darmajaya saat melihatnya di Instagram',
    'Saya pernah mencari tahu tentang IIB Darmajaya karena melihat konten di Instagram. ': 'Pernah mencari tahu tentang IIB Darmajaya karena melihat konten di Instagram',
    'Saya dapat membedakan konten kampus IIB Darmajaya dengan kampus lain'. ': 'Dapat membedakan konten kampus IIB Darmajaya dengan kampus lain',
    'Saya merasa melalui konten Instagram, saya jadi lebih mengenal IIB Darmajaya. ': 'Melalui konten Instagram, jadi lebih mengenal IIB Darmajaya',
    'Menurut Anda, apa yang bisa ditingkatkan dari konten Instagram @darmajayathebest ? ': 'Peningkatan yang dapat dilakukan di konten instagram',
    'Kritik atau saran ': 'masukan',
    'Saya lebih tertarik dengan konten kampus yang berbentuk Reels dari pada gambar biasa atau Feed. ': 'Lebih tertarik dengan konten reels',
    'Sejauh mana Anda merasa mengenal dan mengingat IIB Darmajaya sebagai sebuah kampus tanpa bantuan dari iklan atau promosi? ': 'Mengenal dan mengingat IIB Darmajaya tanpa iklan'
})
df.head().T
```

Gambar 4.3 Kode Program untuk rename

Gambar 4.3 tersebut menampilkan proses penggantian nama kolom pada data kuesioner agar lebih ringkas dan mudah dianalisis. Setiap pertanyaan diubah menjadi label variabel seperti kualitas\_gambar, estetika\_warna, dan konsistensi\_tema. Langkah ini penting untuk mempermudah analisis dan pemodelan data dalam penelitian, khususnya saat membangun klasifikasi dengan algoritma *Naïve Bayes* terhadap persepsi audiens terhadap konten Instagram IIB Darmajaya.

Nama	Daru	sheila febriyanti	Reni Aprilawati	Arya Gunawan	Naina Hibatasya A
Status	Ya	Mahasiswa	Mahasiswa	Mahasiswa	Mahasiswa
kualitas_gambar	5	4	3	5	4
kualitas_gambar_2	5	4	4	5	4
kualitas_gambar_3	5	4	4	5	4
estetika_warna	5	4	3	5	4
estetika_warna_2	5	3	3	5	4
estetika_warna_3	5	4	4	5	4
konsistensi_tema	5	4	4	5	4
konsistensi_tema2	5	3	5	5	4
konsistensi_tema3	5	3	4	5	4
kejelasan_informasi	5	4	4	5	4
kejelasan_informasi2	5	4	4	5	4
kejelasan_informasi3	5	4	4	5	4
estetika_warna_4	5	3	3	5	4
kualitas_gambar_4	5	3	4	5	4
kejelasan_informasi_4	5	3	4	5	4
Lebih tertarik dengan konten reels	5	5	4	5	4
Mengenali konten dari IIB Darmajaya saat melihatnya di Instagram	5	4	4	5	4

Gambar 4. 4 Hasil rename

Gambar 4.4 di atas menunjukkan contoh data hasil kuesioner yang telah mengalami proses penyesuaian nama kategori. Data ini mencakup penilaian dari beberapa responden terhadap berbagai indikator visual pada konten *Instagram* IIB Darmajaya, seperti kualitas gambar, estetika warna, konsistensi tema, hingga kejelasan informasi yang disampaikan.

Setiap responden memberikan skor dari skala Likert (1–5) terhadap masing-masing indikator. Selain itu, terdapat beberapa pertanyaan tambahan berupa pernyataan atau opini terbuka, yang dikategorikan ulang ke dalam tingkat ketertarikan atau persepsi, seperti "Tinggi", "Sedang", atau "Rendah".

## 2. Outlier.

Langkah selanjutnya dalam proses pre-processing adalah melakukan deteksi outlier. Outlier merupakan data yang memiliki nilai ekstrem atau menyimpang jauh dari nilai-nilai lainnya dalam dataset. Hasilnya tidak ditemukan outlier dalam dataset berikut.

```

# Cek outlier
# Menggunakan z-score
def detect_outliers_zscore(df, threshold=3):
    outlier_indices = {}
    for col in df.select_dtypes(include=np.number).columns:
        z_scores = np.abs((df[col] - df[col].mean()) / df[col].std())
        outlier_indices[col] = df.index[z_scores > threshold].tolist()
    return outlier_indices
outliers_zscore = detect_outliers_zscore(df)

print("cek uoutlier dengan z-scoe:")
for col, indices in outliers_zscore.items():
    if indices:
        print(f" {col}: {indices}")
    else:
        print(f" {col}: tidak ada outlier")

# Menggunakan IQR
def detect_outliers_iqr(df, multiplier=1.5):
    outlier_indices = {}
    for col in df.select_dtypes(include=np.number).columns:
        Q1 = df[col].quantile(0.25)
        Q3 = df[col].quantile(0.75)
        IQR = Q3 - Q1
        lower_bound = Q1 - multiplier * IQR
        upper_bound = Q3 + multiplier * IQR
        outlier_indices[col] = df.index[(df[col] < lower_bound) | (df[col] > upper_bound)].tolist()
    return outlier_indices

# IQR 1,5
outliers_iqr = detect_outliers_iqr(df)

# Cek tiap kolom
print("cek outlier dengan IQR:")
for col, indices in outliers_iqr.items():
    if indices:
        print(f" {col}: {indices}")
    else:
        print(f" {col}: tidak ada outlier")

```

Gambar 4. 5 Kode Program mencari Outlier

Gambar 4.5 menampilkan implementasi kode program untuk mendeteksi keberadaan data pencilan (*outlier*) menggunakan metode statistik. Algoritma ini dirancang untuk memindai setiap kolom numerik dalam *dataset* dan mengidentifikasi nilai-nilai ekstrem yang menyimpang jauh dari distribusi data normal. Deteksi ini merupakan langkah krusial untuk menjaga validitas model klasifikasi yang akan dibangun.

```

cek uoutlier dengan z-scoe:
  kualitas_gambar: tidak ada outlier
  kualitas_gambar_2: tidak ada outlier
  kualitas_gambar_3: tidak ada outlier
  estetika_warna: tidak ada outlier
  estetika_warna_2: tidak ada outlier
  estetika_warna_3: tidak ada outlier
  konsistensi_tema: tidak ada outlier
  konsistensi_tema2: tidak ada outlier
  konsistensi_tema3: tidak ada outlier
  kejelasan_informasi: tidak ada outlier
  kejelasan_informasi2: tidak ada outlier
  kejelasan_informasi3: tidak ada outlier
  estetika_warna_4: tidak ada outlier
  kualitas_gambar_4: tidak ada outlier
  kejelasan_informasi_4: tidak ada outlier
Lebih tertarik denga konten reels: tidak ada outlier
Mengenali konten dari IIB Darmajaya saat melihatnya di Instagram: tidak ada outlier
Pernah mencari tahu tentang IIB Darmajaya karena melihat konten di Instagram: tidak ada outlier
Dapat membedakan konten kampus IIB Darmajaya dengan kampus lain: tidak ada outlier
Melalui konten Instagram, jadi lebih mengenal IIB Darmajaya: tidak ada outlier
\cek outlier dengan IQR:
  kualitas_gambar: tidak ada outlier
  kualitas_gambar_2: tidak ada outlier
  kualitas_gambar_3: tidak ada outlier
  estetika_warna: tidak ada outlier
  estetika_warna_2: tidak ada outlier
  estetika_warna_3: tidak ada outlier
  konsistensi_tema: tidak ada outlier
  konsistensi_tema2: tidak ada outlier
  konsistensi_tema3: tidak ada outlier
  kejelasan_informasi: tidak ada outlier
  kejelasan_informasi2: tidak ada outlier
  kejelasan_informasi3: tidak ada outlier
  estetika_warna_4: tidak ada outlier
  kualitas_gambar_4: tidak ada outlier
  kejelasan_informasi_4: tidak ada outlier
Lebih tertarik denga konten reels: tidak ada outlier
Mengenali konten dari IIB Darmajaya saat melihatnya di Instagram: tidak ada outlier
Pernah mencari tahu tentang IIB Darmajaya karena melihat konten di Instagram: tidak ada outlier
Dapat membedakan konten kampus IIB Darmajaya dengan kampus lain: tidak ada outlier
Melalui konten Instagram, jadi lebih mengenal IIB Darmajaya: tidak ada outlier

```

#### Gambar 4. 6 Hasil dari Outlier

Berdasarkan hasil eksekusi program pada Gambar 4.6, *output* teks secara eksplisit mencetak keterangan 'tidak ada outlier' untuk setiap variabel yang diuji, mulai dari kualitas\_gambar hingga kejelasan\_informasi . Hasil ini mengindikasikan bahwa distribusi jawaban responden cenderung seragam dan stabil, sehingga data aman digunakan untuk pelatihan model *Naïve Bayes* tanpa perlu adanya penghapusan data ekstrem.

Dengan tidak ditemukannya outlier, maka dapat disimpulkan bahwa data yang digunakan berada dalam kondisi stabil dan tidak mengandung nilai ekstrem yang dapat mengganggu hasil analisis lebih lanjut. Hal ini memberikan dasar yang kuat untuk tahap berikutnya.

3. Mengubah nilai dari kolom "Mengenal dan mengingat IIB Darmajaya tanpa iklan", supaya proses training dapat dilakukan.

Pada tahap ini dilakukan proses konversi nilai kategorikal menjadi nilai numerik agar data dapat diproses dalam tahap training model machine learning. Salah satu kolom yang diubah adalah kolom "Mengenal dan mengingat IIB Darmajaya tanpa iklan", yang sebelumnya berisi nilai berupa kategori teks seperti:

- Tinggi
- Sedang
- Rendah

Nilai-nilai tersebut tidak dapat langsung digunakan dalam pemodelan karena bersifat non-numerik. Oleh karena itu, dilakukan pengkodean atau encoding untuk mengubah nilai kategorikal tersebut menjadi representasi numerik yang lebih sesuai untuk proses pelatihan model.

Berikut ini adalah skema konversi yang digunakan:

- "Tinggi" → 3
- "Sedang" → 2
- "Rendah" → 1

```
df1['Mengenal dan mengingat IIB Darmajaya tanpa iklan'] = df1['Mengenal dan mengingat IIB Darmajaya tanpa iklan'].replace({  
    'Tinggi': 3,  
    'Sedang': 2,  
    'Rendah': 1  
})  
df1['Mengenal dan mengingat IIB Darmajaya tanpa iklan']
```

Gambar 4. 7 Konversi skala Likert ke numerik.

Gambar 4.7 menunjukkan proses transformasi data kategorikal menjadi data numerik (*encoding*) pada kolom target. Variabel 'Mengenal dan mengingat IIB Darmajaya' yang sebelumnya berisi label teks seperti 'Tinggi', 'Sedang', dan 'Rendah' dikonversi menjadi representasi angka 3, 2, dan 1 . Konversi ini mutlak diperlukan agar algoritma *machine learning* dapat memproses data tersebut dalam perhitungan matematis probabilitas.

Dengan melakukan encoding ini, maka nilai dalam kolom tersebut kini dapat dikenali oleh model klasifikasi atau regresi sebagai nilai numerik yang

merepresentasikan tingkat persepsi responden. Langkah ini juga menjadi bagian penting dari transformasi data agar seluruh fitur dalam dataset memiliki format yang seragam dan sesuai dengan kebutuhan algoritma yang akan digunakan.

4. Menggabungkan kolom sesuai 4 kategori dan membuat rata-ratanya yang kemudian akan dijadikan sebagai variabel X untuk proses training.

Langkah selanjutnya adalah melakukan penggabungan beberapa kolom berdasarkan kategori indikator yang sama, lalu menghitung nilai rata-ratanya. Tujuan dari proses ini adalah untuk menyederhanakan data dan merangkum penilaian dari beberapa indikator menjadi satu nilai representatif untuk masing-masing kategori. Empat kategori utama yang digunakan dalam penelitian ini adalah:

1. Kualitas Gambar
2. Estetika Warna
3. Konsistensi Tema
4. Kejelasan Informasi

Setiap kategori awalnya terdiri dari beberapa sub-atribut, seperti “kualitas\_gambar\_1”, “kualitas\_gambar\_2”, dan seterusnya. Seluruh sub-atribut ini kemudian digabungkan dengan cara menghitung nilai rata-rata dari masing-masing responden untuk setiap kategori tersebut.

Hasil dari penggabungan ini kemudian disimpan dalam kolom baru dengan nama:

- kualitas\_gambar\_gabungan
- kualitas\_gambar\_gabungan
- konsistensi\_tema\_gabungan
- kejelasan\_informasi\_gabungan

```

# menggabungkan kolom "kualitas_gambar", "estetika warna", "konsistensi tema" dan "kejelasan informasi"

#1
df1['kualitas_gambar_gabungan'] = df1[['kualitas_gambar', 'kualitas_gambar_2', 'kualitas_gambar_3','kualitas_gambar_4']].mean(axis=1)
df1.drop(['kualitas_gambar', 'kualitas_gambar_2', 'kualitas_gambar_3','kualitas_gambar_4'], axis=1, inplace=True)
#2
df1['estetika_warna_gabungan'] = df1[['estetika_warna', 'estetika_warna_2', 'estetika_warna_3','estetika_warna_4']].mean(axis=1)
df1.drop(['estetika_warna', 'estetika_warna_2', 'estetika_warna_3','estetika_warna_4'], axis=1, inplace=True)
#3
df1['konsistensi_tema_gabungan'] = df1[['konsistensi_tema', 'konsistensi_tema2', 'konsistensi_tema3']].mean(axis=1)
df1.drop(['konsistensi_tema', 'konsistensi_tema2', 'konsistensi_tema3'], axis=1, inplace=True)
#4
df1['kejelasan_informasi_gabungan'] = df1[['kejelasan_informasi', 'kejelasan_informasi2', 'kejelasan_informasi3','kejelasan_informasi_4']].mean(axis=1)
df1.drop(['kejelasan_informasi', 'kejelasan_informasi2', 'kejelasan_informasi3','kejelasan_informasi_4'], axis=1, inplace=True)

df1.head().T

```

Gambar 4. 8 Kode Program untuk penggabungan beberapa kolom

Gambar 4.8 di atas memperlihatkan implementasi kode program untuk tahap rekayasa fitur (*feature engineering*). Pada potongan kode tersebut, dilakukan proses agregasi data di mana beberapa kolom indikator yang sejenis digabungkan menjadi satu variabel baru. Fungsi `.mean(axis=1)` digunakan untuk menghitung nilai rata-rata dari atribut-atribut seperti `kualitas_gambar`, `estetika_warna`, `konsistensi_tema`, dan `kejelasan_informasi`. Langkah ini bertujuan untuk menyederhanakan dimensi data tanpa mengurangi informasi penting, sehingga menghasilkan variabel independen (X) yang lebih representatif untuk proses pelatihan model *Naïve Bayes*.

<code>kualitas_gambar_gabungan</code>	5.0	3.75	3.75	5.0	4.0
<code>estetika_warna_gabungan</code>	5.0	3.5	3.25	5.0	4.0
<code>konsistensi_tema_gabungan</code>	5.0	3.333333	4.333333	5.0	4.0
<code>kejelasan_informasi_gabungan</code>	5.0	3.75	4.0	5.0	4.0

Gambar 4. 9 Hasil penggabungan nilai rata-rata untuk beberapa kategori

Gambar 4.9 menampilkan tabel hasil transformasi data di mana beberapa atribut telah digabungkan menjadi satu nilai rata-rata. Sebagai contoh, kolom `estetika_warna_gabungan` kini berisi nilai desimal (seperti 5.0, 3.6, 3.25) yang merupakan rata-rata dari sub-pertanyaan terkait warna. Nilai gabungan inilah yang merepresentasikan variabel X secara utuh untuk setiap kategori visual, yang akan menjadi input utama bagi algoritma klasifikasi.

Hasil rata-rata nilai untuk beberapa responden ditampilkan dalam tabel. Pada kolom “`estetika_warna_gabungan`”, nilai rata-rata diperoleh dari penggabungan tiga atribut terkait estetika warna. Hal yang sama juga diterapkan pada kategori lainnya.

Proses ini bertujuan untuk menyederhanakan jumlah fitur yang digunakan dalam pemodelan tanpa menghilangkan informasi penting. Keempat kolom gabungan inilah yang kemudian dijadikan sebagai variabel independen (X) dalam proses pelatihan model machine learning.

## 4.2 Pembahasan

### 4.2.1 Naïve Bayes Training

Dataset akan dibagi untuk digunakan dalam training dan testing dengan rasio 80:20. Training data digunakan untuk membangun dan mengoptimalkan model, sementara Testing data digunakan untuk mengevaluasi seberapa baik model dapat menggeneralisasi data baru yang tidak terlihat selama pelatihan.

```
# Membuat x dan y untuk training dengan y adalah kolom 'Mengenal dan mengingat IIB Darmajaya tanpa iklan'

X = df1.drop(['Nama',
              'Lebih tertarik dengan konten reels',
              'Saya mengetahui IIB Darmajaya sebagai institusi perguruan tinggi. ',
              'Mengenali konten dari IIB Darmajaya saat melihatnya di Instagram',
              'Pernah mencari tahu tentang IIB Darmajaya karena melihat konten di Instagram',
              'Dapat membedakan konten kampus IIB Darmajaya dengan kampus lain',
              'Mengenal dan mengingat IIB Darmajaya tanpa iklan',
              'Melalui konten Instagram, jadi lebih mengenal IIB Darmajaya',
              'Kritik atau saran'],
             axis=1)

y = df1['Mengenal dan mengingat IIB Darmajaya tanpa iklan']
```

Gambar 4. 10 Kode Program untuk menentukan X dan Y

Gambar 4.10 di atas menunjukkan proses pemisahan data menjadi dua bagian utama, yaitu variabel independen (X) dan variabel dependen (y). Variabel X berisi kumpulan kolom yang akan digunakan sebagai input untuk melatih model, sedangkan variabel y berisi kolom target yang ingin diprediksi, yaitu "Mengenal dan mengingat IIB Darmajaya tanpa iklan". Dalam proses ini, beberapa kolom yang tidak relevan seperti kolom nama dan opini responden dihapus dari variabel X agar tidak memengaruhi hasil pelatihan. Pemisahan ini merupakan tahap penting dalam machine learning, di mana model akan dilatih menggunakan data pada X untuk mempelajari hubungan dan pola yang dapat digunakan dalam memprediksi nilai pada y.

```
# model untuk naive bayes
# Bagi data menjadi data training dan testing
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
```

Gambar 4. 11 Kode Program untuk memisahkan variabel X dan Y

Gambar 4.11 di atas menunjukkan proses pembagian data menjadi data pelatihan (training) dan data pengujian (testing) menggunakan fungsi `train_test_split` dari library `sklearn.model_selection`. Pembagian ini dilakukan terhadap variabel independen (X) dan variabel dependen (y), dengan proporsi 80% data digunakan untuk melatih model (`X_train`, `y_train`) dan 20% sisanya untuk menguji performa model (`X_test`, `y_test`). Parameter `random_state=42` digunakan untuk memastikan bahwa proses pembagian data bersifat konsisten setiap kali dijalankan. Proses ini merupakan langkah awal dalam membangun model klasifikasi *Naïve Bayes*, guna memastikan bahwa model dapat dilatih dan diuji secara terpisah sehingga hasil evaluasi lebih objektif dan tidak bias.

```
# Visualisasi sampel yang trained dan tested
import matplotlib.pyplot as plt
# Membuat visualisasi dengan bar chart
labels = ['Data training', 'Data testing']
sizes = [train_samples, test_samples]

plt.figure(figsize=(6, 6))
bars = plt.bar(labels, sizes, color=['skyblue', 'lightcoral'])
plt.ylabel('jumlah')
plt.title('Distribusi data training dan testing')

# Menampilkan angka di atas bar
for bar in bars:
    yval = bar.get_height()
    plt.text(bar.get_x() + bar.get_width()/2, yval, int(yval), va='bottom') # va: vertical alignment

plt.show()
```

Gambar 4. 12 Kode Visual berapa nilai X dan Y

Gambar 4.12 berisi kode program untuk memvisualisasikan proporsi pembagian data *training* dan *testing* ke dalam bentuk diagram batang (*bar chart*). Kode ini memanfaatkan pustaka `matplotlib` untuk membuat grafik yang membandingkan jumlah sampel pada kedua kelompok data, serta menambahkan label angka di atas setiap batang untuk memperjelas informasi jumlah data secara visual.

Untuk memberikan gambaran mengenai proporsi data yang digunakan dalam proses pelatihan dan pengujian model, dilakukan visualisasi distribusi

sampel menggunakan diagram batang. Visualisasi ini menampilkan dua kategori, yaitu data training dan data testing, dengan warna yang dibedakan agar lebih mudah dikenali. Jumlah masing-masing sampel ditampilkan secara langsung di atas batang melalui fungsi anotasi, sehingga informasi yang disampaikan menjadi lebih informatif. Proses ini memanfaatkan library matplotlib, dengan pengaturan ukuran grafik agar tampilan lebih proporsional. Dengan visualisasi ini, pembaca dapat memahami secara jelas bahwa data telah dibagi secara proporsional untuk menjamin keakuratan dalam evaluasi performa model.

Setelah proses pembagian data dilakukan, langkah selanjutnya adalah membangun model klasifikasi menggunakan algoritma *Naïve Bayes*, tepatnya dengan pendekatan Gaussian *Naïve Bayes*. Model diinisialisasi menggunakan fungsi GaussianNB() dari pustaka sklearn. *Naïve Bayes* kemudian dilatih menggunakan data training yang telah disiapkan sebelumnya melalui metode .fit(X\_train, y\_train). Tahap pelatihan ini bertujuan agar model dapat mempelajari pola dari data input yang ada serta hubungannya terhadap output. Setelah model selesai dilatih, proses dilanjutkan dengan melakukan prediksi terhadap data testing menggunakan metode .predict(X\_test). Hasil dari prediksi ini akan digunakan untuk mengevaluasi performa model dalam memprediksi kemampuan pengguna mengenal dan mengingat IIB Darmajaya tanpa iklan.

```
# Inisialisasi model Naive Bayes (Gaussian Naive Bayes)
model = GaussianNB()

# Latih model menggunakan data training
model.fit(X_train, y_train)

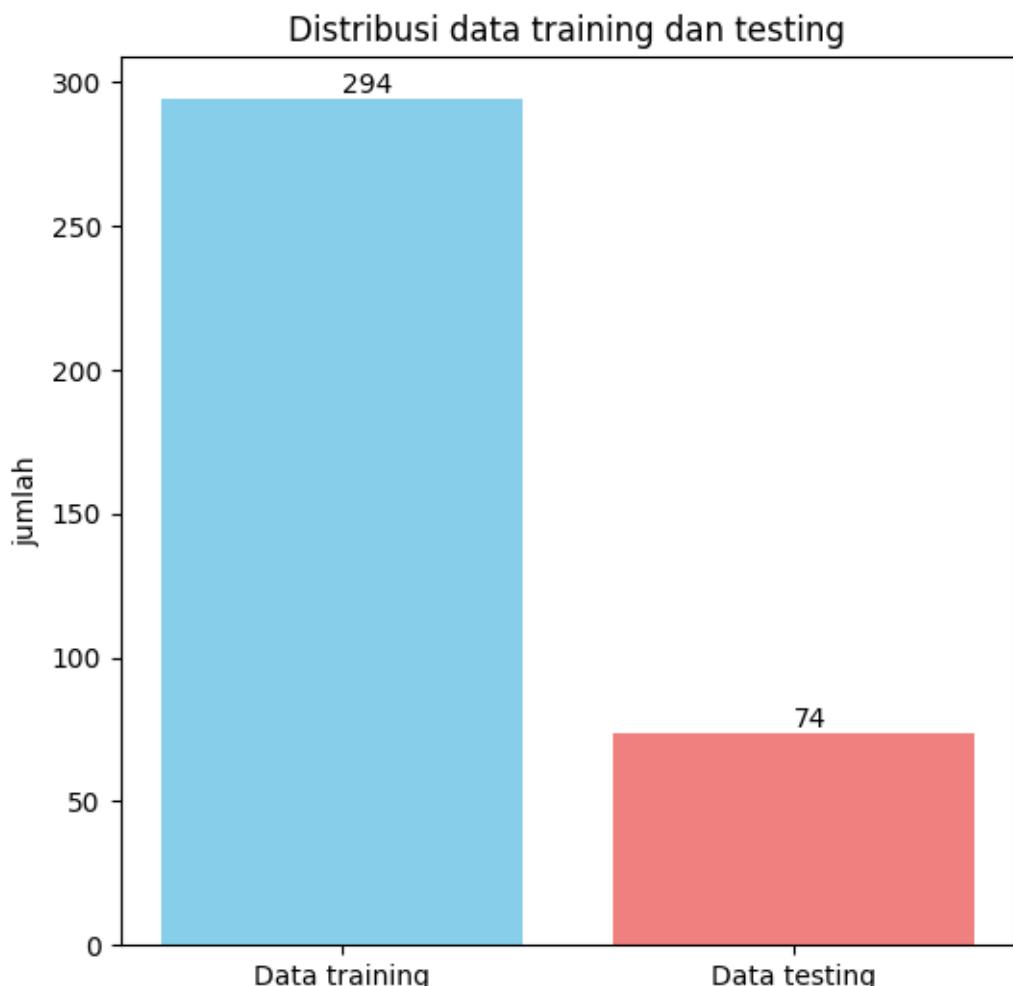
# Lakukan prediksi pada data testing
y_pred = model.predict(X_test)
```

Gambar 4. 13 Kode Program training dan testing data

Gambar 4.13 menampilkan kode program inti untuk proses pelatihan model. Algoritma GaussianNB() diinisialisasi dan dilatih menggunakan data X\_train dan y\_train melalui fungsi .fit(). Setelah model mempelajari pola dari data

latih, fungsi `.predict()` digunakan pada data `X_test` untuk menghasilkan prediksi, yang nantinya akan dibandingkan dengan data aktual untuk evaluasi.

Distribusi data yang telah dibagi ditampilkan dalam bentuk diagram batang yang memperlihatkan jumlah sampel pada masing-masing bagian. Berdasarkan hasil visualisasi, diketahui bahwa sebanyak **294 data** digunakan sebagai data training, sementara **74 data** sisanya digunakan sebagai data testing. Pembagian ini mengikuti proporsi 80:20, yang merupakan standar umum dalam proses pelatihan model machine learning untuk menjaga keseimbangan antara kemampuan model dalam belajar dan menghindari overfitting. Warna batang yang berbeda digunakan untuk membedakan kedua kelompok data, yaitu *skyblue* untuk data training dan *lightcoral* untuk data testing. Jumlah masing-masing data juga ditampilkan di atas batang agar informasi kuantitatif lebih mudah terbaca secara langsung oleh pembaca.



Gambar 4. 14 Hasil visualisasi data training dan testing

Berdasarkan visualisasi pada Gambar 4.14, terlihat distribusi jumlah sampel data yang telah dibagi. Batang berwarna biru menunjukkan **Data Training sebanyak 294 sampel**, sedangkan batang berwarna merah menunjukkan **Data Testing sebanyak 74 sampel**. Visualisasi ini menegaskan bahwa proporsi data sudah seimbang dan memadai untuk proses pelatihan serta validasi model *Naïve Bayes*.

- **Training** data digunakan untuk membangun dan mengoptimalkan model, sementara
- **Testing** data digunakan untuk mengevaluasi seberapa baik model dapat menggeneralisasi data baru yang tidak terlihat selama pelatihan.

#### 4.2.2 Implementasi Teori *Naïve Bayes*

Dalam proses klasifikasi *Brand awareness* menggunakan algoritma Gaussian *Naïve Bayes*, sistem bekerja berdasarkan prinsip teorema *Bayes* yang dirumuskan sebagai:

$$P(H|D) = \frac{P(D|H) \cdot P(H)}{P(D)}$$

Namun karena nilai  $P(D)$  bersifat konstan untuk semua kelas, maka cukup dihitung berdasarkan:

$$P(C|X) \propto P(C) \cdot \prod_{i=1}^n P(x_i|C)$$

- $C$ : kelas (Tinggi, Sedang, Rendah)
- $x_1$  : fitur i (kualitas gambar, estetika warna, dst.)
- $X = (x_1, x_2, \dots, x_n)$ : fitur gabungan
- $P(C)$ : prior probability
- $P(x_i | C)$ : likelihood

##### 1. Persiapan Dataset

Dataset mencakup:

Tabel 4. 1 Dataset

Responden	Kualitas Gambar	Estetika Warna	Konsistensi Tema	Kejelasan Informasi	<i>Brand awareness</i>
R1	4.2	4.5	4.0	4.3	Tinggi
R2	3.0	3.2	3.0	3.1	Sedang
R3	2.0	2.1	2.3	2.0	Rendah
...	...	...	...	...	...

##### 2. Perhitungan Prior Probability

Jumlah total responden adalah **368**, dengan distribusi sebagai berikut:

- Tinggi: 150 →  $P(\text{Tinggi}) = 150/368$

- Sedang: 140 →  $P(\text{Sedang})=140/368$
- Rendah: 78 →  $P(\text{Rendah})=78/368$

Nilai prior ini menunjukkan seberapa besar kemungkinan suatu data termasuk dalam masing-masing kelas sebelum mempertimbangkan fitur-fitur lainnya.

### 3. Perhitungan Likelihood (*Gaussian Naïve Bayes*) $P(x_i | C)$

Untuk fitur bertipe numerik (misalnya skala Likert), digunakan rumus Gaussian:

$$P(x_i | C) = \frac{1}{\sqrt{2\pi \cdot \sigma_C^2}} \cdot e^{-\frac{(x_i - \mu_C)^2}{2 \cdot \sigma_C^2}}$$

Sebagai contoh, jika ingin menghitung likelihood untuk fitur **Estetika Warna** pada kelas **Tinggi**, dengan parameter:

- Rata-rata ( $\mu$ ) = 4.3
- Standar deviasi ( $\sigma$ ) = 0.4
- Nilai input (x) = 4.0

Maka:

$$P(x = 4.0 | C = \text{Tinggi}) = \frac{1}{\sqrt{2\pi \cdot 0.4^2}} \cdot e^{-\frac{(4.0 - 4.3)^2}{2 \cdot 0.4^2}}$$

Hasil dari perhitungan ini dilakukan untuk setiap fitur yang tersedia terhadap masing-masing kelas.

### 4. Menghitung Posterior Probability

Setelah mendapatkan seluruh nilai likelihood untuk setiap fitur terhadap masing-masing kelas, langkah berikutnya adalah mengalikan semuanya dengan nilai prior dari masing-masing kelas:

$$P(C|X) = P(C) \cdot P(x_1|C) \cdot P(x_2|C) \cdot \dots \cdot P(x_n|C)$$

Hasil dari proses ini akan berupa nilai probabilitas akhir untuk setiap kelas (Tinggi, Sedang, Rendah). Kelas dengan **nilai tertinggi** akan dijadikan **prediksi akhir** untuk data input X.

### **4.2.3 Evaluasi Model**

Setelah proses pelatihan model *Naïve Bayes* selesai dilakukan, langkah selanjutnya adalah melakukan evaluasi untuk mengukur seberapa baik model dapat melakukan prediksi terhadap data yang belum pernah dilihat sebelumnya. Evaluasi ini bertujuan untuk menilai performa model dalam konteks klasifikasi, serta memastikan bahwa model tidak hanya bekerja baik pada data latih, tetapi juga memiliki kemampuan generalisasi yang baik pada data uji (testing data).

Pada evaluasi ini digunakan beberapa metrik pengukuran utama, yaitu:

- Accuracy: Mengukur proporsi prediksi yang benar dari keseluruhan data uji.
- Precision: Mengukur tingkat ketepatan prediksi positif yang dilakukan model.
- Recall: Mengukur sejauh mana model mampu menangkap seluruh data positif yang sebenarnya.
- F1-Score: Merupakan rata-rata dari precision dan recall, digunakan ketika diperlukan keseimbangan antara keduanya.

```

# visualisasi hasil dari akurasi, precision, recall dan f1-score

import matplotlib.pyplot as plt
# Mendapatkan classification report sebagai dictionary
report = classification_report(y_test, y_pred, output_dict=True)

# Ekstrak nilai akurasi, precision, recall, dan f1-score (rata-rata tertimbang)
accuracy = report['accuracy']
precision = report['weighted avg']['precision']
recall = report['weighted avg']['recall']
f1_score = report['weighted avg']['f1-score']

# Siapkan data untuk visualisasi
metrics = ['Accuracy', 'Precision', 'Recall', 'F1-Score']
values = [accuracy, precision, recall, f1_score]

# Buat bar plot
plt.figure(figsize=(8, 6))
sns.barplot(x=metrics, y=values)
plt.ylim(0, 1.1) # Batasi sumbu y antara 0 dan 1
plt.title('Evaluasi Model Naive Bayes')
plt.ylabel('Score')

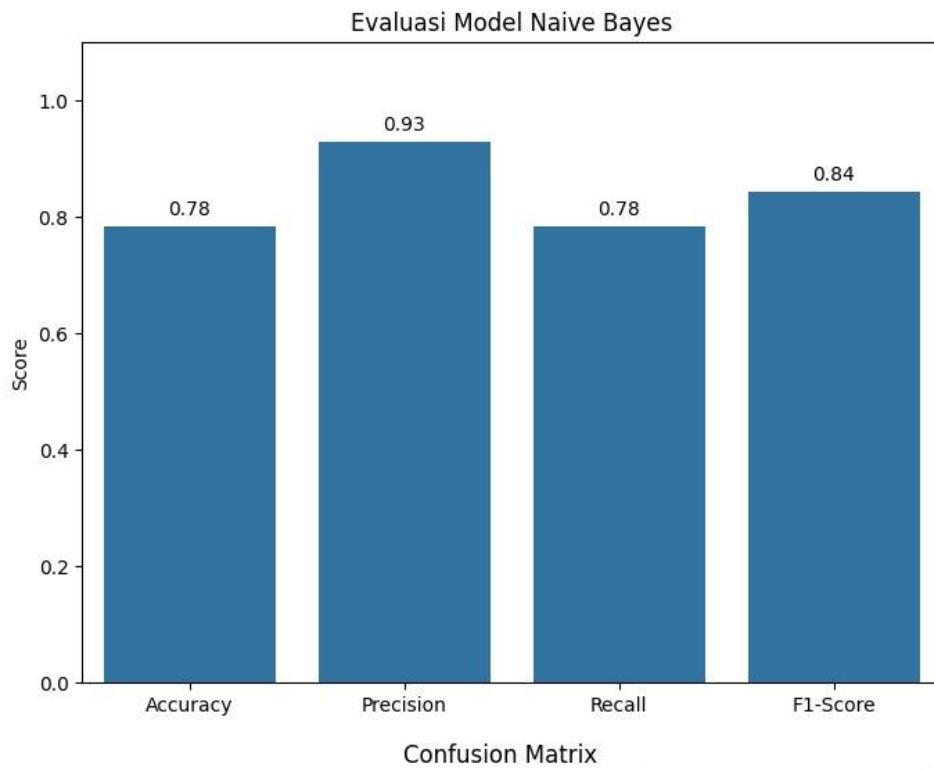
# Tambahkan nilai di atas bar
for i, value in enumerate(values):
    plt.text(i, value + 0.02, f'{value:.2f}', ha='center')

plt.show()

```

Gambar 4. 15 Kode visualisasi dari nilai akurasi, presisi, *recall* & *F1-score*

Gambar 4.15 menampilkan kode program yang dirancang untuk menghitung dan memvisualisasikan metrik evaluasi model. Kode ini mengekstrak nilai-nilai utama seperti Akurasi, Presisi, *Recall*, dan *F1-Score* dari laporan klasifikasi (*classification report*), kemudian menyajikannya dalam bentuk diagram batang agar performa model dapat dianalisis dengan lebih mudah dan intuitif.



Gambar 4. 16 Hasil Evaluasi Model

Grafik 4.16 di atas menunjukkan bahwa model *Naïve Bayes* yang telah dibangun memiliki performa yang cukup baik secara keseluruhan. Nilai accuracy sebesar 78% menunjukkan bahwa 78 dari setiap 100 prediksi yang dilakukan model adalah benar.

Yang menarik adalah nilai precision yang sangat tinggi, yaitu 0.93, yang berarti bahwa model sangat akurat dalam memprediksi kelas positif. Namun, nilai recall masih setara dengan accuracy yaitu 0.78, yang menandakan masih ada beberapa data positif yang tidak berhasil dikenali oleh model.

Sementara itu, nilai F1-score sebesar 0.84 mengindikasikan bahwa model mampu menjaga keseimbangan antara precision dan recall, yang penting terutama dalam kasus di mana kesalahan prediksi dapat berdampak besar. Secara keseluruhan, model ini sudah cukup layak untuk digunakan.

Pada gambar dibawah ini fitur yang paling memengaruhi *brand awareness* terutama tentang mengenali dan mengingat Darmajaya tanpa iklan. Yang paling diperhatikan adalah tentang estetika warnanya, dan kejelasan informasi.

```
result = permutation_importance(model, X_test, y_test, n_repeats=10, random_state=42)
importance_scores = pd.Series(result.importances_mean, index=X.columns).sort_values(ascending=False)

print("Top Influential Features:")
print(importance_scores.head())

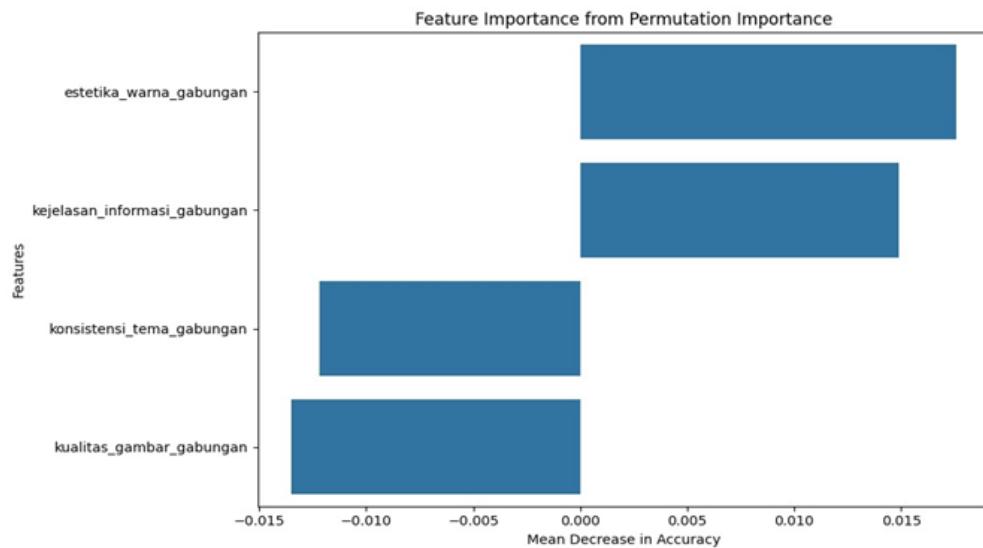
Top Influential Features:
estetika_warna_gabungan      0.017568
kejelasan_informasi_gabungan  0.014865
konsistensi_tema_gabungan    -0.012162
kualitas_gambar_gabungan     -0.013514
dtype: float64

# visualisasi dari variabel importance_scores

plt.figure(figsize=(10, 6))
sns.barplot(x=importance_scores, y=importance_scores.index)
plt.title('Feature Importance from Permutation Importance')
plt.xlabel('Mean Decrease in Accuracy')
plt.ylabel('Features')
plt.tight_layout()
plt.show()
```

Gambar 4. 17 Kode Program Feature Importances Permutation

Gambar 4.17 memperlihatkan kode program untuk menghitung *Feature Importance* menggunakan metode *Permutation Importance*. Kode ini bertujuan untuk mengukur seberapa besar pengaruh setiap variabel (seperti estetika warna atau kejelasan informasi) terhadap akurasi model dalam memprediksi *brand awareness*, dengan cara mengacak nilai fitur dan mengamati penurunan performa model.



Gambar 4. 18 Feature Importances Permutation

Dari hasil visualisasi pada gambar 4.18 tersebut, terlihat bahwa "estetika\_warna\_gabungan" merupakan faktor paling dominan. Ini berarti bahwa tampilan visual, seperti pemilihan warna yang menarik dan harmonis, sangat berperan dalam membentuk persepsi dan daya ingat audiens terhadap merek Darmajaya. Warna yang tepat mampu membangun citra yang kuat di benak pengguna.

Selanjutnya, "kejelasan\_informasi\_gabungan" juga memiliki pengaruh besar. Artinya, konten yang disampaikan melalui media sosial atau platform visual lainnya harus mudah dipahami dan informatif. Informasi yang jelas membuat audiens lebih cepat mengenali dan mengingat Darmajaya.

Fitur lain seperti "konsistensi\_tema\_gabungan" dan "kualitas\_gambar\_gabungan" juga turut berkontribusi, meskipun pengaruhnya tidak sebesar dua faktor sebelumnya. Konsistensi tema menjaga identitas visual agar tidak membingungkan audiens, sementara kualitas gambar mendukung daya tarik visual secara keseluruhan.

Dengan demikian, dapat disimpulkan bahwa dalam membangun *brand awareness* secara organik, unsur visual seperti estetika warna dan kejelasan pesan menjadi kunci utama yang harus diperhatikan dalam setiap konten yang dipublikasikan.

Gambar di bawah ini memperlihatkan analisis feature importance dari model klasifikasi yang digunakan untuk mengukur tingkat *brand awareness* terhadap IIB Darmajaya melalui konten *Instagram*. Hasilnya menunjukkan bahwa fitur "kejelasan\_informasi\_gabungan" memberikan kontribusi paling besar terhadap akurasi model. Artinya, semakin jelas dan mudah dipahami informasi yang disampaikan dalam konten *Instagram*, semakin besar kemungkinan audiens mengenali dan mengingat Darmajaya.

```

result = permutation_importance(model, X1_test, y1_test, n_repeats=10, random_state=42)
importance_scores = pd.Series(result.importances_mean, index=X1.columns).sort_values(ascending=False)

print("Top Influential Features:")
print(importance_scores.head())

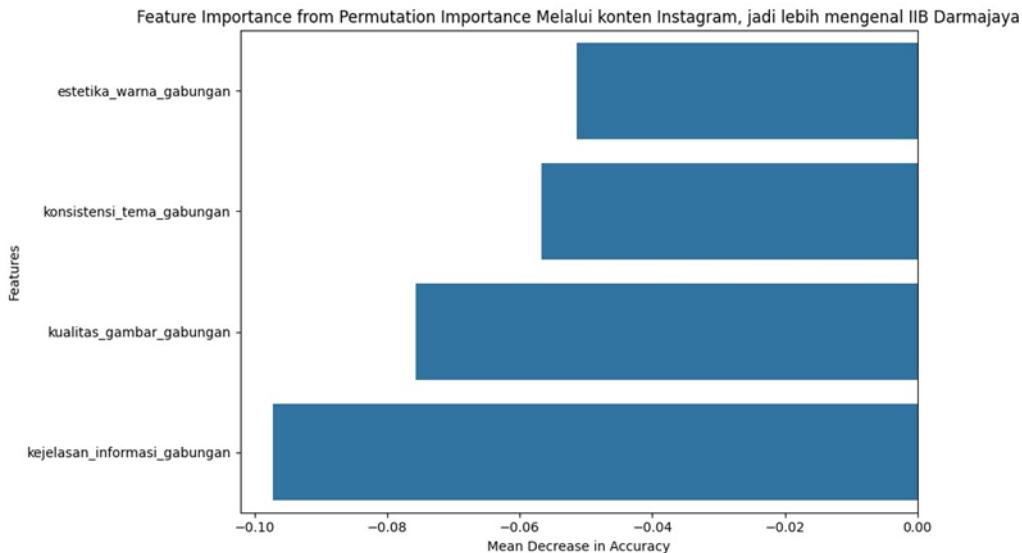
Top Influential Features:
estetika_warna_gabungan      -0.051351
konsistensi_tema_gabungan     -0.056757
kualitas_gambar_gabungan      -0.075676
kejelasan_informasi_gabungan  -0.097297
dtype: float64

# Visualisasi feature importance
plt.figure(figsize=(10, 6))
sns.barplot(x=importance_scores, y=importance_scores.index)
plt.title('Feature Importance from Permutation Importance Melalui konten Instagram, jadi lebih mengenal IIB Darmajaya')
plt.xlabel('Mean Decrease in Accuracy')
plt.ylabel('Features')
plt.tight_layout()
plt.show()

```

Gambar 4. 19 Kode Program Feature importance model Naïve Bayes.

Gambar 4.19 menampilkan kode program lanjutan untuk menganalisis fitur *importance* pada skenario spesifik, yaitu variabel 'jadi lebih mengenal IIB Darmajaya'. Analisis ini dilakukan untuk melihat apakah faktor yang memengaruhi pemahaman mendalam audiens berbeda dengan faktor yang memengaruhi *awareness* umum.



Gambar 4. 20 Hasil Permutation Importance Fitur Konten Instagram

Pada Gambar 4.20 diatas terdapat temuan ini menegaskan bahwa konten yang informatif dan komunikatif lebih efektif dalam membangun awareness dibandingkan sekadar tampilan visual yang menarik. Meskipun elemen seperti "estetika\_warna\_gabungan" secara umum dianggap penting dalam aspek branding, dalam konteks ini justru menunjukkan kontribusi yang rendah atau bahkan negatif secara statistik. Hal ini mengindikasikan bahwa aspek visual tidak selalu menjamin keberhasilan dalam menyampaikan pesan atau membentuk kesadaran merek.

Kesimpulannya, dalam membangun *brand awareness* melalui media sosial seperti *Instagram*, fokus utama sebaiknya diberikan pada bagaimana menyampaikan pesan yang jelas dan mudah dipahami audiens. Visual yang menarik memang mendukung, tetapi tidak cukup jika tidak diimbangi dengan konten yang komunikatif dan bermakna.

Gambar di bawah ini menunjukkan hasil analisis feature importance dari model klasifikasi yang digunakan untuk melihat faktor-faktor yang memengaruhi kemampuan audiens dalam membedakan konten kampus IIB Darmajaya dengan konten dari kampus lain melalui platform *Instagram*.

```

result = permutation_importance(model, X2_test, y2_test, n_repeats=10, random_state=42)
importance_scores = pd.Series(result.importances_mean, index=X1.columns).sort_values(ascending=False)

print("Top Influential Features:")
print(importance_scores.head())

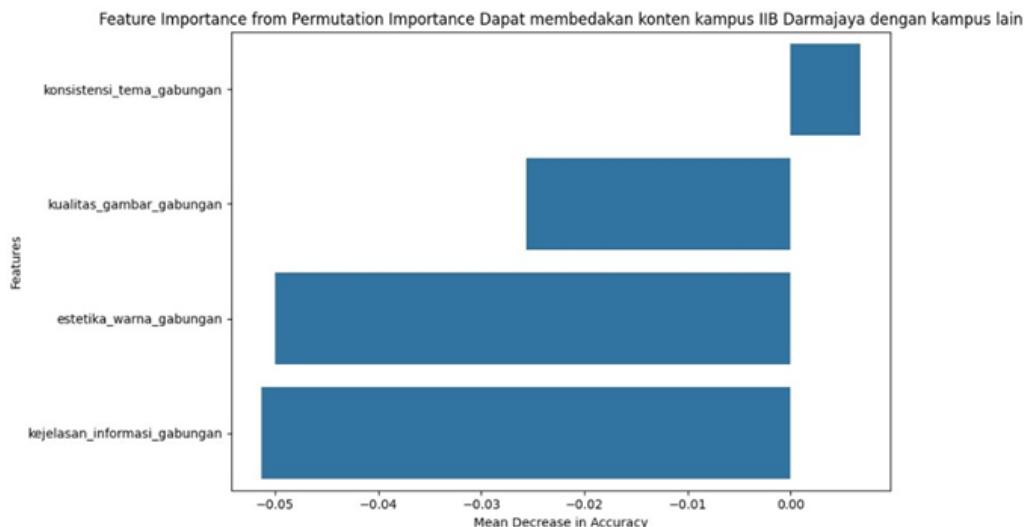
Top Influential Features:
kualitas_gambar_gabungan           -0.004054
Dapat membedakan konten kampus IIB Darmajaya dengan kampus lain -0.006757
estetika_warna_gabungan            -0.024324
Melalui konten Instagram, jadi lebih mengenal IIB Darmajaya      -0.043243
konsistensi_tema_gabungan          -0.074324
dtype: float64

# Visualisasi feature importance
plt.figure(figsize=(10, 6))
sns.barplot(x=importance_scores, y=importance_scores.index)
plt.title('Feature Importance from Permutation Importance Dapat membedakan konten kampus IIB Darmajaya dengan kampus lain')
plt.xlabel('Mean Decrease in Accuracy')
plt.ylabel('Features')
plt.tight_layout()
plt.show()

```

Gambar 4. 21 Kode Fitur pembeda konten IIB Darmajaya dari kampus lain.

Gambar 4.21 berisi kode program untuk menganalisis fitur yang paling berpengaruh dalam membedakan konten IIB Darmajaya dengan kampus lain. Fokus analisis ini adalah untuk mengidentifikasi elemen visual mana yang menjadi ciri khas atau identitas pembeda (*differentiator*) di mata audiens Instagram.



Gambar 4. 22 Hasil Permutation Importance Fitur Konten Pembeda Kampus

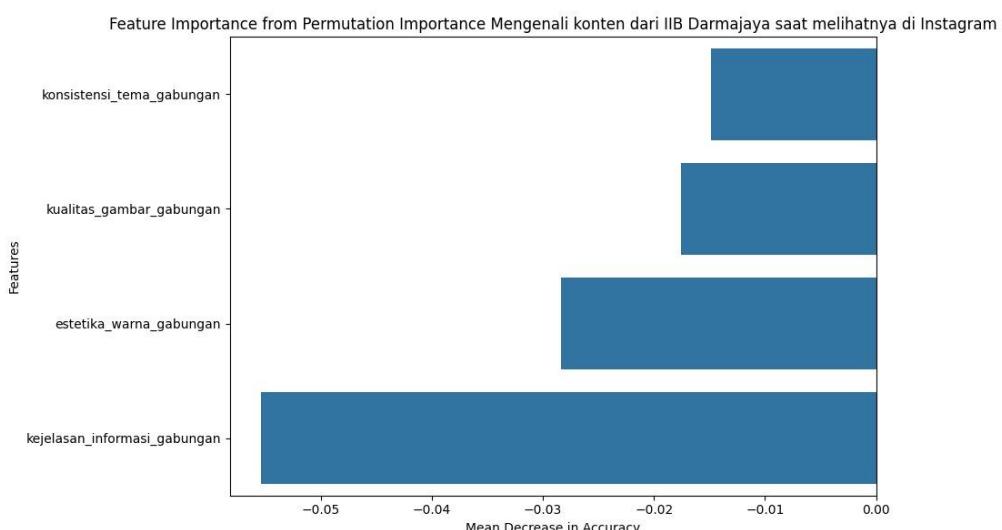
Pada Gambar 4.22 diatas hasilnya memperlihatkan bahwa fitur "konsistensi\_tema\_gabungan" menjadi aspek yang paling berkontribusi terhadap akurasi model. Ini menunjukkan bahwa penggunaan tema visual yang konsisten seperti layout, tone warna, gaya desain, dan pola penyampaian konten memiliki peran besar dalam menciptakan identitas yang khas dan mudah dikenali. Dengan

kata lain, audiens dapat lebih mudah mengidentifikasi bahwa sebuah konten berasal dari Darmajaya hanya dengan melihat keselarasan tampilannya.

Sebaliknya, fitur "kejelasan\_informasi\_gabungan" menunjukkan pengaruh yang lebih rendah dalam konteks pembedaan antar kampus. Ini bisa jadi karena kejelasan informasi lebih berkaitan dengan pemahaman isi, bukan dengan tampilan visual yang menjadi fokus utama ketika audiens membedakan asal konten di media sosial.

Dengan demikian, dapat disimpulkan bahwa konsistensi visual dan warna khas merupakan kunci utama dalam membangun identitas konten kampus yang mudah dikenali dan dibedakan dari institusi lain. Hal ini menjadi strategi penting bagi Darmajaya dalam memperkuat positioning-nya di ranah digital, khususnya di platform visual seperti *Instagram*.

Gambar di bawah ini menampilkan hasil analisis feature importance untuk mengidentifikasi fitur-fitur utama yang berkontribusi dalam proses audiens mengenali bahwa sebuah konten berasal dari akun *Instagram* resmi IIB Darmajaya.



Gambar 4. 23 Kode Feature importance Konten Instagram IIB Darmajaya.

Dari hasil visualisasi pada gambar 4.23 tersebut, tampak bahwa "konsistensi\_tema\_gabungan" merupakan faktor paling dominan. Hal ini menunjukkan bahwa konten yang selalu menggunakan pola desain yang seragam

baik dari segi layout, tone warna, hingga gaya penyampaian mampu menciptakan ciri khas visual yang kuat. Dengan begitu, audiens dapat langsung mengenali bahwa konten tersebut berasal dari Darmajaya hanya dengan melihat tampilannya sekilas.

Selain itu, "kualitas\_gambar\_gabungan" juga memiliki pengaruh yang signifikan. Kualitas visual yang baik seperti gambar yang tajam, tidak buram, dan memiliki pencahayaan yang tepat meningkatkan daya tarik konten. "estetika\_warna\_gabungan" juga menunjukkan kontribusi yang cukup penting. Warna yang konsisten dan mudah dikenali memperkuat identitas visual kampus, sehingga semakin mempertegas bahwa konten tersebut berasal dari Darmajaya. "kejelasan\_informasi\_gabungan" justru menunjukkan pengaruh yang negatif dalam konteks ini. Ini bisa dijelaskan karena proses mengenali sumber konten di *Instagram* cenderung lebih mengandalkan aspek visual dibandingkan isi teks atau narasi. Artinya, meskipun informasi yang jelas penting secara umum, dalam konteks pengenalan visual cepat, pengguna lebih dipengaruhi oleh elemen desain.

Secara keseluruhan, temuan ini memperkuat pentingnya membangun identitas visual yang kuat dan konsisten, serta menjaga kualitas visual agar konten dapat dengan mudah dikenali sebagai bagian dari brand IIB Darmajaya di tengah banyaknya informasi yang berseliweran di media sosial.

#### 4.2.4 Interpretasi Hasil

Gambar di bawah ini merupakan hasil visualisasi *word cloud* yang menggambarkan kata-kata yang paling sering muncul dalam tanggapan responden terkait peningkatan konten *Instagram* IIB Darmajaya.

```
# wordcloud dari kolom "Peningkatan yang dapat dilakukan di konten instagram"
# Combining all text data from the specified column
text = " ".join(review for review in df['Peningkatan yang dapat dilakukan di konten instagram'].astype(str))

# Define custom stopwords
custom_stopwords = set(STOPWORDS)
custom_stopwords.update(['kadang', 'di', 'nggak', 'ga', 'nya', 'ada', 'terlalu', 'cepat', 'ke', 'kurang', 'dan', 'lebih', 'yang', 'atau', 'jadi', 'Beberapa', 'beberapa', 'terasa', 'kayak', 'simple'])

# Generate the word cloud
wordcloud = WordCloud(stopwords=custom_stopwords, background_color="white").generate(text)

# Display the generated image:
plt.figure(figsize=(10, 8))
plt.imshow(wordcloud, interpolation='bilinear')
plt.axis('off')
plt.show()
```

Gambar 4. 24 Kode Program untuk visual “Peningkatan yang dapat dilakukan”

Gambar 4.24 memperlihatkan kode program untuk menghasilkan visualisasi *Word Cloud* dari kolom saran responden. Kode ini memproses teks

masukan pada kolom 'Peningkatan yang dapat dilakukan', menghapus kata hubung (*stopwords*), dan menyusun kata-kata yang paling sering muncul menjadi representasi visual berbentuk awan kata.



Gambar 4. 25 Word Cloud Kata Kunci Peningkatan Konten Instagram

Dari hasil yang terdapat pada gambar 4.25 tersebut, terlihat bahwa kata “konten”, “reels”, dan “video” mendominasi frekuensi kemunculan. Ini menunjukkan bahwa sebagian besar masukan atau kritik pengguna berkaitan erat dengan format dan daya tarik konten yang disajikan.

Selain itu, kata “menarik”, “unik”, dan “bosan” juga muncul cukup menonjol, yang menunjukkan adanya kebutuhan akan inovasi dan penyegaran dalam gaya penyampaian konten. Responden menekankan bahwa konten yang disajikan saat ini masih kurang variatif atau belum cukup menarik, sehingga disarankan untuk memperkuat unsur kreativitas, storytelling, dan visual yang engaging.

Kemunculan kata seperti “**fokus**”, “**feed**”, “**panjang**”, “**caption**”, serta “**musik**” juga menunjukkan bahwa elemen teknis dan presentasi konten turut diperhatikan audiens. Mereka tidak hanya mengomentari isi, tetapi juga format penyajiannya seperti durasi video, pemilihan audio latar, dan kekuatan narasi dalam caption.