

elyluasom

November 3, 2025

```
[98]: # =====  
# TUGAS-A-1-EKSPLORASI ANALISIS SENTIMEN  
# NAMA: ALFA RENALDO ALUSKA  
# NRP: 5026221144  
# =====
```

## 1 1: Praproses data

Untuk nomor 1. Sudah dilakukan di kelas Ibu Irma, nanti hanya upload file output saja.

#2: Baseline Model (BoW+NB) Untuk nomor 2: Bag of Words, Naive Bayes.

```
[99]: # === BAGIAN 2: BASELINE MODEL ===  
  
import pandas as pd  
from sklearn.model_selection import train_test_split  
from sklearn.feature_extraction.text import CountVectorizer  
from sklearn.naive_bayes import MultinomialNB  
from sklearn.pipeline import Pipeline  
from sklearn.metrics import classification_report, accuracy_score  
from collections import Counter  
  
# Load dataset  
df = pd.read_csv("7-garudaindonesia_news_cleaned_simple.csv")  
df.columns = df.columns.str.strip()  
  
# Pilih kolom teks dan label  
X = df['konten_normalized']  
y = df['sentiment']  
  
# Split data  
X_train, X_test, y_train, y_test = train_test_split(X, y, stratify=y,   
↳test_size=0.2, random_state=42)  
  
# Buat pipeline BoW + NB  
pipeline_bow_nb_asli = Pipeline([  
    ('vect', CountVectorizer()),
```

```

        ('clf', MultinomialNB())
    ])

    # Latih model baseline
    pipeline_bow_nb_asli.fit(X_train, y_train)
    y_pred_baseline = pipeline_bow_nb_asli.predict(X_test)

    # Evaluasi lengkap
    print("\n=== BASELINE MODEL (BoW + NB) ===")
    acc_bow_nb_asli = accuracy_score(y_test, y_pred_baseline)
    report_bow_nb_asli = classification_report(y_test, y_pred_baseline,
        ↪ zero_division=0, output_dict=True)

    print(f"Akurasi: {acc_bow_nb_asli:.4f}\n")
    print("F1-Score per Kelas:")
    for label in ['Negative', 'Neutral', 'Positive']:
        f1 = report_bow_nb_asli[label]['f1-score']
        precision = report_bow_nb_asli[label]['precision']
        recall = report_bow_nb_asli[label]['recall']
        support = report_bow_nb_asli[label]['support']
        print(f"- {label:<8} | Precision: {precision:.2f} | Recall: {recall:.2f} | ↪
        ↪ F1: {f1:.2f} | Support: {support}")

    print(f"\nMacro Avg F1-Score    : {report_bow_nb_asli['macro avg']['f1-score']:.
        ↪ 4f}")
    print(f"\nWeighted Avg F1-Score: {report_bow_nb_asli['weighted avg']['f1-score']:.
        ↪ 4f}")

```

```

=== BASELINE MODEL (BoW + NB) ===
Akurasi: 0.8404

```

F1-Score per Kelas:

```

- Negative | Precision: 0.82 | Recall: 0.86 | F1: 0.84 | Support: 37.0
- Neutral  | Precision: 0.75 | Recall: 0.53 | F1: 0.62 | Support: 17.0
- Positive | Precision: 0.88 | Recall: 0.95 | F1: 0.92 | Support: 40.0

```

Macro Avg F1-Score : 0.7928

Weighted Avg F1-Score: 0.8334

#3: Augmentasi Indobert Context embedding.

[100]: # === BAGIAN 3: AUGMENTASI INDOBERT ===

```

!pip install -q nlpaug transformers Sastrawi
import nlpaug.augmenter.word as naw
#from Sastrawi.Stemmer.StemmerFactory import StemmerFactory
from tqdm import tqdm

```

```

# Inisialisasi IndoBERT
model_name = 'indobenchmark/indobert-base-p1'
try:
    augmenter = naw.ContextualWordEmbsAug(
        model_path=model_name,
        action="substitute",
        device='cuda'
    )
    print(" IndoBERT berhasil dimuat di GPU.")
except RuntimeError:
    augmenter = naw.ContextualWordEmbsAug(
        model_path=model_name,
        action="substitute",
        device='cpu'
    )
    print(" IndoBERT fallback ke CPU.")

# Inisialisasi stemmer
#stemmer = StemmerFactory().create_stemmer()

# Siapkan data untuk augmentasi
neutral_indices = y_train[y_train == 'Neutral'].index
negative_indices = y_train[y_train == 'Negative'].index

neutral_texts_to_aug = df.loc[neutral_indices]['konten_normalized'].tolist()
negative_texts_to_aug = df.loc[negative_indices]['konten_normalized'].tolist()

n_positive = Counter(y_train)['Positive']
n_neutral = Counter(y_train)['Neutral']
n_negative = Counter(y_train)['Negative']

# n_aug_neutral = max(1, round((n_positive - n_neutral) / n_neutral))
#n_aug_negative = max(1, round((n_positive - n_negative) / n_negative))
target_per_class = Counter(y_train)['Positive'] # kita samakan ke jumlah
↳Positive
#n_aug_neutral = max(0, round((target_per_class - n_neutral) / n_neutral))
#n_aug_negative = max(0, round((target_per_class - n_negative) / n_negative))
n_aug_neutral = target_per_class - n_neutral # jumlah teks tambahan yang
↳dibutuhkan
n_aug_negative = target_per_class - n_negative

print(f"\nTarget distribusi per kelas: {target_per_class}")
print(f"Jumlah data sebelum augmentasi:")
print(f"- Neutral : {n_neutral}")
print(f"- Negative : {n_negative}")

```

```

print(f"- Positive : {n_positive}")
print(f"\nTarget Augmentasi:")
print(f"- Neutral → {n_aug_neutral}x")
print(f"- Negative → {n_aug_negative}x")

# Augmentasi dengan progress bar
#print("\nAugmentasi teks Neutral...")
#new_neutral_texts = [augmenter.augment(text, n=n_aug_neutral) for text in
    ↳ tqdm(neutral_texts_to_aug, desc='Augment Neutral')]
#new_neutral_texts = sum(new_neutral_texts, [])
print("\nAugmentasi teks Neutral...")
new_neutral_texts = []
i = 0
while len(new_neutral_texts) < n_aug_neutral:
    text = neutral_texts_to_aug[i % len(neutral_texts_to_aug)]
    new_neutral_texts.extend(augmenter.augment(text, n=1))
    i += 1
#for text in tqdm(neutral_texts_to_aug, desc='Augment Neutral'):
#    new_neutral_texts.extend(augmenter.augment(text, n=1))
#    if len(new_neutral_texts) >= n_aug_neutral:
#        break

#print("Augmentasi teks Negative...")
#new_negative_texts = [augmenter.augment(text, n=n_aug_negative) for text in
    ↳ tqdm(negative_texts_to_aug, desc='Augment Negative')]
#new_negative_texts = sum(new_negative_texts, [])
print("Augmentasi teks Negative...")
new_negative_texts = []
i = 0
while len(new_negative_texts) < n_aug_negative:
    text = negative_texts_to_aug[i % len(negative_texts_to_aug)]
    new_negative_texts.extend(augmenter.augment(text, n=1))
    i += 1

# Stemming dengan progress bar
#stemmed_new_neutral = [stemmer.stem(text) for text in tqdm(new_neutral_texts,
    ↳ desc='Stemming Neutral')]
#stemmed_new_negative = [stemmer.stem(text) for text in
    ↳ tqdm(new_negative_texts, desc='Stemming Negative')]

#new_neutral_labels = ['Neutral'] * len(stemmed_new_neutral)
#new_negative_labels = ['Negative'] * len(stemmed_new_negative)

# Gabungkan data latihan

```

```

#X_train_aug_bert = pd.concat([X_train.reset_index(drop=True), pd.
    ↳Series(stemmed_new_neutral), pd.Series(stemmed_new_negative)],
    ↳ignore_index=True)
#y_train_aug_bert = pd.concat([y_train.reset_index(drop=True), pd.
    ↳Series(new_neutral_labels), pd.Series(new_negative_labels)],
    ↳ignore_index=True)

new_neutral_labels = ['Neutral'] * len(new_neutral_texts)
new_negative_labels = ['Negative'] * len(new_negative_texts)

# Gabungkan data latih
X_train_aug_bert = pd.concat([X_train.reset_index(drop=True), pd.
    ↳Series(new_neutral_texts), pd.Series(new_negative_texts)], ignore_index=True)
y_train_aug_bert = pd.concat([y_train.reset_index(drop=True), pd.
    ↳Series(new_neutral_labels), pd.Series(new_negative_labels)],
    ↳ignore_index=True)

print(f"\nDistribusi Setelah Augmentasi: {Counter(y_train_aug_bert)}")

```

IndoBERT berhasil dimuat di GPU.

Target distribusi per kelas: 161

Jumlah data sebelum augmentasi:

- Neutral : 68
- Negative : 146
- Positive : 161

Target Augmentasi:

- Neutral → 93x
- Negative → 15x

Augmentasi teks Neutral...

Augmentasi teks Negative...

Distribusi Setelah Augmentasi: Counter({'Neutral': 161, 'Positive': 161, 'Negative': 161})

#4: Train Augmented BoW+NB

```

[101]: # Latih ulang model dengan data augmented
pipeline_bow_nb_aug.fit(X_train_aug_bert, y_train_aug_bert)
y_pred_aug_bert = pipeline_bow_nb_aug.predict(X_test)

# Evaluasi model augmented
print("\n=== MODEL SETELAH AUGMENTASI (IndoBERT) ===")
acc_aug_bert = accuracy_score(y_test, y_pred_aug_bert)

```

```

report_aug_bert = classification_report(y_test, y_pred_aug_bert,
    ↪zero_division=0, output_dict=True)

print(f"Akurasi: {acc_aug_bert:.4f}\n")
print("F1-Score per Kelas:")
for label in ['Negative', 'Neutral', 'Positive']:
    f1 = report_aug_bert[label]['f1-score']
    precision = report_aug_bert[label]['precision']
    recall = report_aug_bert[label]['recall']
    support = report_aug_bert[label]['support']
    print(f"- {label:<8} | Precision: {precision:.2f} | Recall: {recall:.2f} |
    ↪F1: {f1:.2f} | Support: {support}")

print(f"\nMacro Avg F1-Score    : {report_aug_bert['macro avg']['f1-score']:.
    ↪4f}")
print(f"Weighted Avg F1-Score: {report_aug_bert['weighted avg']['f1-score']:.
    ↪4f}")

# Perbandingan akhir
f1_baseline = report_bow_nb_asli['Neutral']['f1-score']
f1_augmented = report_aug_bert['Neutral']['f1-score']

print("\n=== PERBANDINGAN F1-SCORE (Neutral) ===")
print(f"1. Baseline (Asli):           {f1_baseline:.4f}")
print(f"2. IndoBERT Augmentation:       {f1_augmented:.4f}")

# Simpan hasil ke CSV
df_augmented_bert = pd.DataFrame({
    'final_text': X_train_aug_bert,
    'score_label': y_train_aug_bert
})
df_augmented_bert.to_csv("garuda_augmented_bert.csv", index=False)
print(f"\n Data augmented disimpan sebagai 'garuda_augmented_bert.csv'
    ↪({len(df_augmented_bert)} baris)")

```

=== MODEL SETELAH AUGMENTASI (IndoBERT) ===  
 Akurasi: 0.8617

F1-Score per Kelas:

- Negative | Precision: 0.86 | Recall: 0.84 | F1: 0.85 | Support: 37.0
- Neutral | Precision: 0.75 | Recall: 0.71 | F1: 0.73 | Support: 17.0
- Positive | Precision: 0.90 | Recall: 0.95 | F1: 0.93 | Support: 40.0

Macro Avg F1-Score : 0.8345

Weighted Avg F1-Score: 0.8602

=== PERBANDINGAN F1-SCORE (Neutral) ===

1. Baseline (Asli):	0.6207
2. IndoBERT Augmentation:	0.7273

Data augmented disimpan sebagai 'garuda\_augmented\_bert.csv' (483 baris)

#5: Hasil & Banding BoW+NB

```
[102]: import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.metrics import confusion_matrix
from collections import Counter

labels = ['Negative', 'Neutral', 'Positive']

# === CEK ISI DATASET HASIL AUGMENTASI ===
print("\n=== CEK DATA LATIH SETELAH AUGMENTASI ===")
print(f"Jumlah data latih sebelum augmentasi: {len(X_train)}")
print(f"Jumlah data latih setelah augmentasi: {len(X_train_aug_bert)}")

print("\nDistribusi label sebelum augmentasi:")
print(Counter(y_train))

print("\nDistribusi label setelah augmentasi:")
print(Counter(y_train_aug_bert))

#print("\nContoh hasil augmentasi Neutral:")
#for i in range(3):
#    print("-", stemmed_new_neutral[i])

print("\nContoh hasil augmentasi Neutral:")
for i in range(3):
    print("-", new_neutral_texts[i])

#print("\nContoh hasil augmentasi Negative:")
#for i in range(3):
#    print("-", stemmed_new_negative[i])

print("\nContoh hasil augmentasi Negative:")
for i in range(3):
    print("-", new_negative_texts[i])

# === BASELINE: STATISTIKA DESKRIPTIF ===
print("\n=== BASELINE: STATISTIKA DESKRIPTIF ===")
print("Distribusi label di seluruh dataset:")
print(Counter(df['sentiment']))
```

```

print("\nDistribusi Label Asli (y_test):")
print(Counter(y_test))

print("\nDistribusi Prediksi (Baseline):")
print(Counter([str(label) for label in y_pred_baseline]))

cm_base = confusion_matrix(y_test, y_pred_baseline, labels=labels)
plt.figure(figsize=(6,5))
sns.heatmap(cm_base, annot=True, fmt='d', cmap='Blues', xticklabels=labels,
    ↳ yticklabels=labels)
plt.title("Confusion Matrix - Baseline")
plt.xlabel("Predicted")
plt.ylabel("Actual")
plt.show()

print("\nMetrik per Kelas (Baseline):")
for label in labels:
    p = report_bow_nb_asli[label]['precision']
    r = report_bow_nb_asli[label]['recall']
    f1 = report_bow_nb_asli[label]['f1-score']
    s = report_bow_nb_asli[label]['support']
    print(f"- {label:<8} | Precision: {p:.2f} | Recall: {r:.2f} | F1: {f1:.2f} |
    ↳ Support: {s}")

print(f"\nMacro Avg F1-Score : {report_bow_nb_asli['macro avg']['f1-score']:.
    ↳ 4f}")
print(f"Weighted Avg F1-Score: {report_bow_nb_asli['weighted avg']['f1-score']:.
    ↳ 4f}")

# === AUGMENTED: STATISTIKA DESKRIPTIF ===
print("\n=== AUGMENTED: STATISTIKA DESKRIPTIF ===")
print("Distribusi label setelah augmentasi:")
print(Counter(y_train_aug_bert))

print("\nDistribusi Label Asli (y_test):")
print(Counter(y_test))

print("\nDistribusi Prediksi (Augmented):")
print(Counter([str(label) for label in y_pred_aug_bert]))

cm_aug = confusion_matrix(y_test, y_pred_aug_bert, labels=labels)
plt.figure(figsize=(6,5))
sns.heatmap(cm_aug, annot=True, fmt='d', cmap='Greens', xticklabels=labels,
    ↳ yticklabels=labels)
plt.title("Confusion Matrix - IndoBERT Augmented")

```



```

plt.xlabel("Predicted")
plt.ylabel("Actual")
plt.show()

print("\nMetrik per Kelas (Augmented):")
for label in labels:
    p = report_aug_bert[label]['precision']
    r = report_aug_bert[label]['recall']
    f1 = report_aug_bert[label]['f1-score']
    s = report_aug_bert[label]['support']
    print(f"- {label:<8} | Precision: {p:.2f} | Recall: {r:.2f} | F1: {f1:.2f} | Support: {s}")

print(f"\nMacro Avg F1-Score : {report_aug_bert['macro avg']['f1-score']:.4f}")
print(f"Weighted Avg F1-Score: {report_aug_bert['weighted avg']['f1-score']:.4f}")

```

=== CEK DATA LATIH SETELAH AUGMENTASI ===

Jumlah data latih sebelum augmentasi: 375

Jumlah data latih setelah augmentasi: 483

Distribusi label sebelum augmentasi:

Counter({'Positive': 161, 'Negative': 146, 'Neutral': 68})

Distribusi label setelah augmentasi:

Counter({'Neutral': 161, 'Positive': 161, 'Negative': 161})

Contoh hasil augmentasi Neutral:

- tempo cowok jakarta presiden amerika serikat as donald trump telah mengumumkan komitmen indonesia untuk membeli cita boeing trump mengatakan rencana itu sebagai bagian dari kesepakatan dagang kedua negara setelah dirinya berdiskusi langsung dengan presiden prabowo subianto baca berita dengan sedikit iklan klik di sini indonesia telah berkomitmen untuk tindak inspir energi as senilai us miliar produk pertanian amerika senilai us miliar dan jet boeing banyak di antaranya adalah seri tulis trump melalui unggahan di media sosial truth social pada selasa juli waktu as scroll ke bawah untuk melanjutkan membaca baca berita dengan sedikit iklan klik di sini menanggapi hal itu presiden prabowo mengungkapkan pembelian puluhan armada pesawat boeing diperlukan untuk pengembangan pt garuda indonesia jari tbk dia bertekad membesarkan maskapai pelat merah yang lahir saat tanah perang kemerdekaan ri tersebut jadi garuda harus menjadi lambang indonesia kita bertekad saya bertekad untuk membesarkan garuda kata prabowo usai tiba dari kunjungan kerja ke beberapa negara di pangkalan tni angkatan udara lanud halim perdanakusuma jakarta rabu juli lantas seperti apa perkembangan pembelian pesawat boeing garuda klaim ada yang minat mendanai vice cukup vp corporate secretary garuda indonesia cahyadi indrananto

mengatakan perseroan secara paralel sedang berkomunikasi dengan sejumlah pihak pemberi dana potensial menurut dia beberapa pihak telah menunjukkan minat mereka untuk mendanai pembelian puluhan pesawat itu baca juga boeing proyeksikan faktor butuh unit pesawat baru ada pihak pihak yang telah menunjukkan minat untuk menyediakan pendanaan ucap cahyadi ketika dihubungi tempo pada senin juli akan tetapi dia enggan merinci pihak pihak potensial yang dimaksud seluruhnya lanjut dia tengah berproses dan belum mencapai kesepakatan akhir karena ini work ini progress saya belum dapat menyampaikan detailnya namun semoga prosesnya lancar dan nanti akan segera kami sampaikan di kesempatan pertama ujar cahyadi dia turut pinjaman dana sebesar rp triliun dari badan pengelola investasi daya anagata nusantara bpi danantara pada juni lalu tidak bisa dimanfaatkan untuk membeli pesawat boeing pasalnya menurut dia dana pinjaman itu dialokasikan untuk perawatan hingga perbaikan pesawat shareholder loan dari danantara bulan juni lalu adalah hal yang berbeda kolaborasi garuda dengan danantara saat itu adalah untuk mendanai kebutuhan maintenance repair and overhaul mro pemeliharaan perbaikan dan pemeriksaan baik garuda maupun citilink kata cahyadi sedang mencari pinjaman sementara itu pada kesempatan terpisah direktur utama garuda indonesia wamildan tsani panjaitan mengatakan bahwa perusahaan kini sedang menjalin komunikasi dengan beberapa pihak untuk menghimpun dana terutama terkait rencana pembelian pesawat boeing kini was indonesia menggunakan dana pinjaman dari danantara sebesar rp triliun dan dana internal dia menjelaskan bahwa pembelian puluhan armada pesawat itu sejalan dengan upaya penyehatan keuangan sebagaimana toko dalam rancangan restrukturisasi alokasi dana tersebut lanjut dia telah disetujui oleh menteri badan usaha milik negara bumn erick thohir dan rapat umum pemegang saham luar biasa rupslb pada senin juni lalu selain itu perseroan juga secara paralel tengah menjalin komunikasi dengan sejumlah pihak pemberi dana potensial ucap wamildan dalam keterbukaan informasi bursa efek indonesia bei pada senin juli terkait progres pembelian pesawat boeing dia mengungkapkan bahwa garuda indonesia masih terus berkomunikasi dengan pabrikan asal as tersebut pembahasannya mencakup rincian kebutuhan armada yang sesuai dengan pangsa pasar garuda indonesia perseroan dan boeing tengah melakukan komunikasi secara intensif ucap wamildan dia menegaskan pembelian armada pesawat menjadi bagian dari strategi jangka panjang penyehatan keuangan garuda indonesia sekaligus transformasi bisnis melalui penguatan armada dan optimalisasi jaringan penerbangan dalam lima tahun ke depan masih dalam tahap negosiasi pada kesempatan yang lain menteri koordinator menko bidang perekonomian airlangga hartarto mengatakan rencana pembelian pesawat boeing oleh garuda indonesia masih dalam tahap negosiasi antara kedua pihak dia menyebut kedua pihak masih belum mencapai kesepakatan pesawat garuda itu sedang dalam proses negosiasi business tapi business b b antara boeing dengan garuda jadi nanti teknisnya kita tunggu perkembangan selanjutnya ujar airlangga usai sosialisasi tarif trump bersama asosiasi pengusaha di kantor kemenko bidang perekonomian jakarta senin juli seperti dikutip dari antara annisa febiola adil al hasan dan ilona estherina berkontribusi dalam penulisan artikel ini

- garuda indonesia bantah berikan tiket gratis seumur hidup untuk bayi yang lahir di pesawat suara com seorang warganet baru baru ini membagikan pengalaman hidupnya yang mendapat hak istimewa dari maskapai penerbangan garuda ia mengaku bisa naik pesawat garuda seumur hidupnya selera membayar sepeser pun bukan tanpa

sebab rupanya warganet yang diketahui bernama john ini lahir di pesawat garuda pada silam hal ini dibagikannya riset unggahan akun tiktok mjohngaruda pada juli dalam unggahan tersebut ia bercerita bahwa dirinya lahir ketika ibunya sedang berada di dalam pesawat garuda gaskesehatan warganet naik pesawat garuda gratis tiktok mjohngaruda alhasil ia mendapat semacam hadiah dari maskapai tersebut 69 semen naik pesawat 89 secara contoh lahir di pesawat garuda dalam keadaan prematur dan dapat reward naik pesawat kemanapun gratis tanpa biaya mikroorganisme pun dan bisa menyandang nama besar seperti garuda tulisnya tak hanya itu pria yang lahir pada desember tersebut juga diberi nama lengkap muhammad john garuda putra meski mendapat hak istimewa namun john menyebut terdapat limit pemakaian hasi ia tidak bisa sesuka hati naik pesawat garuda unggahan yang telah ditonton sebanyak lebih dari juta penayangan itu pun menuai beragam komentar dari publik besok kalau sudah nikah suruh lahiran di bank ah biar bisa ambil uang sepuasnya tulis akun ham baca juga jadi proyek percontohan ribu siswa di tangerang bakal dapat makan siang gratis terima kasih insightnya aku jadi kepikiran besok mau lahiran di pesawat juga komentar rim privilegenya enggak main main tambah bill pihak garuda indonesia lantas merespon kabar viral tersebut dari pesan singkat tersebut akun instagram garuda indonesia menjelaskan pihaknya belum ada memberikan apresiasi berupa tiket penerbangan gratis ke bayi yang lahir di pesawat selamat siang kak hingga saat ini garuda indonesia belum ada kebijakan tuh apresiasi berupa tiket penerbangan gratis untuk bayi yang lahir di pesawat terima kasih tegasnya

- penulisan com jakarta garuda indonesia telah menyelesaikan proses investigasi internal terkait ct kehilangan barang milik salah satu penumpang berupa iphone pada penerbangan enggak rute jakarta melbourne tanggal juni direktur niaga garuda indonesia adek r susardi mengatakan investigasi internal garuda indonesia penyeb secara menyeluruh melalui koordinasi intensif dengan berbagai lahir terkait yang dilakukan pada juni berdasarkan hasil investigasi belum terdapat bukti untuk mengindikasikan adanya keterlibatan penggalan berpangkat garuda indonesia dalam dugaan kehilangan barang dimaksud ujar adek dalam pernyataan tertulis jumat advertisement sembarang bentuk komitmen untuk memastikan keamanan dan kenyamanan penumpang garuda indonesia telah melaporkan kejadian ini secara resmi kepada bareskrim polri selanjutnya pihak maskapai akan mendukung penuh proses penyelidikan pandang dan siap menyediakan informasi yang diperlukan kepolisian demi tercapainya penyelesaian proses hukum secara adil obyektif dan transparan bagi seluruh pihak kami juga terus melakukan komunikasi dengan penumpang yang bersangkutan imbuhnya laporan penumpang pada kejadian ini menjadi bahan evaluasi bagi prosedur operasional dan pelayanan garuda indonesia sebagai bagian dari upaya kami memberikan service excellence here penumpang pungkas berani

Contoh hasil augmentasi Negative:

- viral penumpang garuda indonesia merokok di pesawat terancam kena denda rp miliar bantal merokok di pesawat diatur dalam undang undang dengan sanksi berat demi keselamatan dan kenyamanan penumpang sebuah asli yang menunjukkan seorang penumpang kelas bisnis garuda indonesia kedapatan menggunakan rokok elektrik di tangkis pesawat viral di media sosial kejadian tersebut terjadi pada penerbangan enggak rute jakarta medan kualanamu direktur utama garuda indonesia wamildan

tsani menegaskan bahwa pihaknya telah mengambil tindakan tegas terhadap penumpang yang melanggar aturan tersebut wamildan menjelaskan bahwa awak pesawat telah mengikuti prosedur yang berlaku dengan memberikan teguran verbal dua kali kepada penumpang yang bersangkutan mengacu pada ketentuan terkait penumpang yang mengganggu disruptive passenger adapun penumpang bersangkutan setibanya mayoritas bandara kualanamu langsung dijemput oleh tim avsec untuk pelaksanaan prosedur investigasi lebih lanjut kata wamildan dalam keterangannya sabtu merokok di dalam pesawat udara di indonesia dilarang tegas dan diatur dalam undang undang nomor tahun tentang penerbangan khususnya pasal ayat dan pasal larangan ini sejalan dengan standar internasional dari organisasi penerbangan sipil internasional icao yang telah diadopsi oleh banyak negara di dunia pelanggaran terhadap larangan ini dapat berakibat sangat serius baik bagi pelanggar itu sendiri maupun bagi sebaya semua penumpang dan awak kabin dalam konteks ini sanksi yang diterapkan bagi mereka yang melanggar start karakter di pesawat sangatlah berat tercermin maksimal yang dapat dikenakan mencapai rp miliar dan pelanggar dapat dijatuhi hukuman penjara selama maksimal tahun ini menunjukkan betapa seriusnya pemerintah dalam menegakkan aturan demi keselamatan penerbangan larangan merokok di pesawat bukan tanpa alasan terdapat beberapa faktor penting yang mendasari keputusan melalui termasuk keselamatan kesehatan dan kenyamanan penumpang lainnya faktor keselamatan dalam larangan merokok keselamatan adalah alasan utama di balik larangan merokok di pesawat rokok merupakan sumber potensial kebakaran terutama di lingkungan kabin pesawat yang kering dalam situasi darurat memadamkan kebakaran di dalam pesawat sangat sulit dan dapat membahayakan keselamatan semua penumpang serta awak kabin oleh karena itu larangan ini diambil untuk mencegah kemungkinan terjadinya insiden yang dapat mengancam nyawa selain itu sirkulasi udara di dalam pesawat juga menjadi perhatian residu asap rokok dapat mengotori saluran udara mengganggu kinerja filter dan menurunkan efisiensi sistem sirkulasi udara hal ini dapat menyebabkan kualitas udara di dalam kabin menurun yang berdampak negatif pada kesehatan penumpang dampak kesehatan bagi penumpang asap rokok mengandung banyak bahan kimia berbahaya yang dapat membahayakan kesehatan penumpang lain terutama bagi mereka yang memiliki masalah pernapasan seperti asma atau bronkitis paparan asap rokok dapat menyebabkan iritasi pada saluran pernapasan dan memperburuk kondisi kesehatan yang sudah ada oleh karena itu larangan ini juga bertujuan untuk melindungi kesehatan semua penumpang di dalam pesawat di samping itu kenyamanan penumpang juga menjadi pertimbangan penting asap rokok dapat membahayakan kenyamanan penumpang lain yang tidak merokok karena bau dan iritasi yang ditimbulkannya larangan merokok di pesawat menciptakan lingkungan yang lebih nyaman bagi semua orang yang berada di dalam kabin aturan dan regulasi terkait rokok elektrik berdasarkan surat edaran se djpu penumpang diperkenankan membawa maksimal satu rokok elektrik dalam kondisi tertentu seperti baterai terlepas dan kapasitas baterai maksimal wh cairan isi ulang rokok elektrik yang dibawa pun dibatasi hingga ml dan harus dikemas dalam kantong plastik meskipun rokok elektrik boleh dibawa ke dalam pesawat penggunaannya tetap dilarang garuda indonesia sebagai salah satu maskapai penerbangan terbesar di indonesia menegaskan komitmennya untuk selalu mengutamakan keselamatan keamanan dan kenyamanan penerbangan sesuai regulasi yang berlaku pihak maskapai mengungkapkan penyesalan atas insiden pelanggaran larangan merokok dan menekankan bahwa merokok termasuk penggunaan

rokok elektrik di dalam kabin pesawat merupakan pelanggaran serius yang tidak ditoleransi

- ambon Kompas esensial pesawat Garuda Indonesia dengan nomor penerbangan GIA tujuan Jakarta. Ambon gagal mendarat di Bandar Udara Pattimura Ambon Sabtu sore penerbangan terpaksa dialihkan ke Bandara Sri Lurah akibat angin kencang dan hujan deras yang melanda wilayah. Sifatnya hanya satu penerbangan sifat-sifat batal divert tapi kendari KDI karena tailwind kecepatan angin knot kata Rahmat Sugeng Widianand, Legal Compliance and Stakeholder Relation Department Head PT Angkasa Pura I, cabang Bandara Pattimura kepada Kompas.com. Baca juga iPhone Raib di Enggak Garuda. Sebut awak kabin tak terbukti terlibat. Gandeng Bareskrim kecepatan angin melebihi batas aman. Rahmat menjelaskan bahwa cuaca di Kota Ambon selama dua hari terakhir 224 ekstrem dengan hujan tanpa henti dan angin kencang. Puncaknya terjadi pada Sabtu sore ketika kecepatan angin mencapai knot melebihi batas aman knot untuk pendaratan pesawat Garuda seharusnya mendarat pukul 17.00 tapi karena cuaca buruk pesawat dialihkan ke Kendari. Saat ini pesawat belum kembali ke Ambon. Ujar Rahmat, Bandara tetap siaga meski hanya satu penerbangan yang terdampak. Pihak Bandara Pattimura tetap siaga menghadapi kemungkinan cuaca ekstrem. Lanjutkan informasi keberangkatan selanjutnya masih menunggu perkembangan kondisi cuaca.

- Alfabeta Concern Jakarta viral di media sosial pesawat Garuda percaya penerbangan Surabaya menuju Bandara Soekarno Hatta dikabarkan tak bisa terbang akibat mesin parts mati. Manajemen Garuda Indonesia pun buka terasasi dalam pesan di media sosial. Seorang penumpang membagikan pengalamannya. Cuma sempat panik di atas udara karena kabarnya mesin pesawat mati dan harus putar balik ke Bandara asal yakni Surabaya. Corporate Secretary RW Head Garuda Indonesia Cahyadi Indrananto menjelaskan kejadian tersebut terjadi pada Rabu malam September penerbangan enggak dengan rute Surabaya Jakarta. Itu sempat lepas landas dari Bandara Internasional Juanda Surabaya pada pukul 17.00 WIB. Advertisement tapi tak lama kemudian pesawat harus melakukan return tapi base RTB dan mendarat kembali di Surabaya pada pukul 18.00 WIB. Garuda Indonesia memohon maaf atas ketidaknyamanan yang dialami penumpang dalam muncul tersebut dan memastikan seluruh penumpang telah mendapatkan service recovery sesuai dengan aturan yang berlaku. Ujar Cahyadi dalam keterangan tertulis Jumat, pesawat berjenis Boeing 737 MAX 8 dengan registrasi D-AMGB tersebut turut telah menjalani prosedur pre-flight dan dinyatakan laik untuk terbang.

=== BASELINE: STATISTIKA DESKRIPTIF ===

Distribusi label di seluruh dataset:

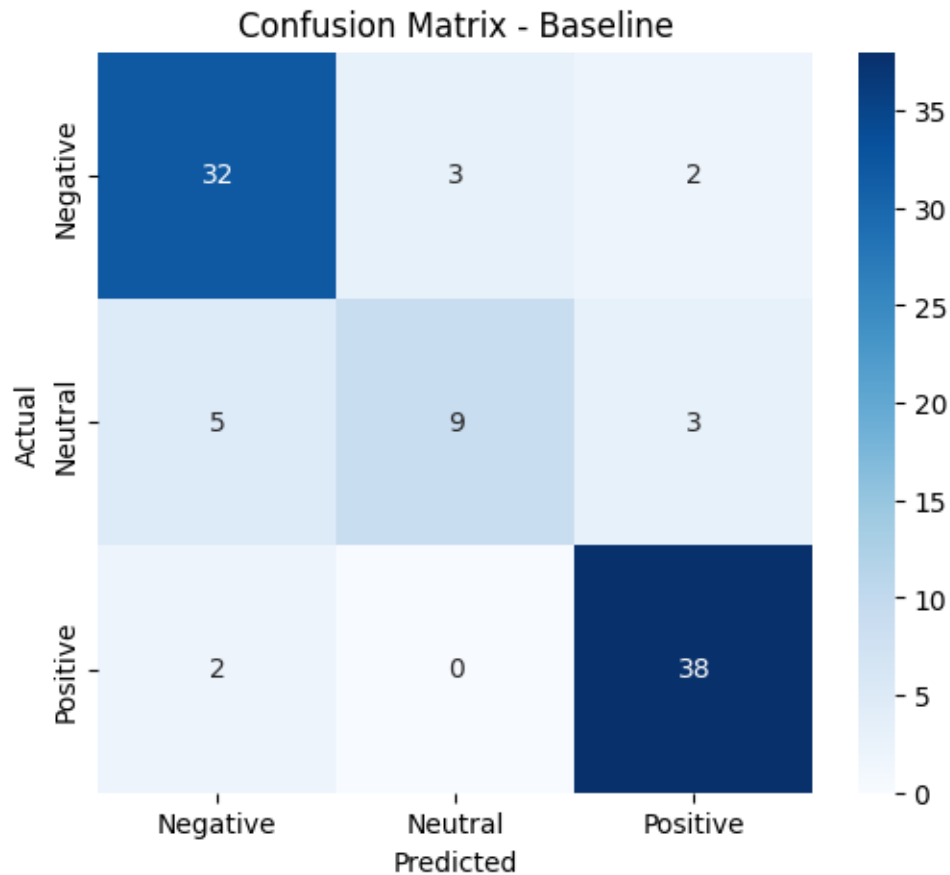
```
Counter({'Positive': 201, 'Negative': 183, 'Neutral': 85})
```

Distribusi Label Asli (y\_test):

```
Counter({'Positive': 40, 'Negative': 37, 'Neutral': 17})
```

Distribusi Prediksi (Baseline):

```
Counter({'Positive': 43, 'Negative': 39, 'Neutral': 12})
```



Metrik per Kelas (Baseline):

- Negative | Precision: 0.82 | Recall: 0.86 | F1: 0.84 | Support: 37.0
- Neutral | Precision: 0.75 | Recall: 0.53 | F1: 0.62 | Support: 17.0
- Positive | Precision: 0.88 | Recall: 0.95 | F1: 0.92 | Support: 40.0

Macro Avg F1-Score : 0.7928

Weighted Avg F1-Score: 0.8334

=== AUGMENTED: STATISTIKA DESKRIPTIF ===

Distribusi label setelah augmentasi:

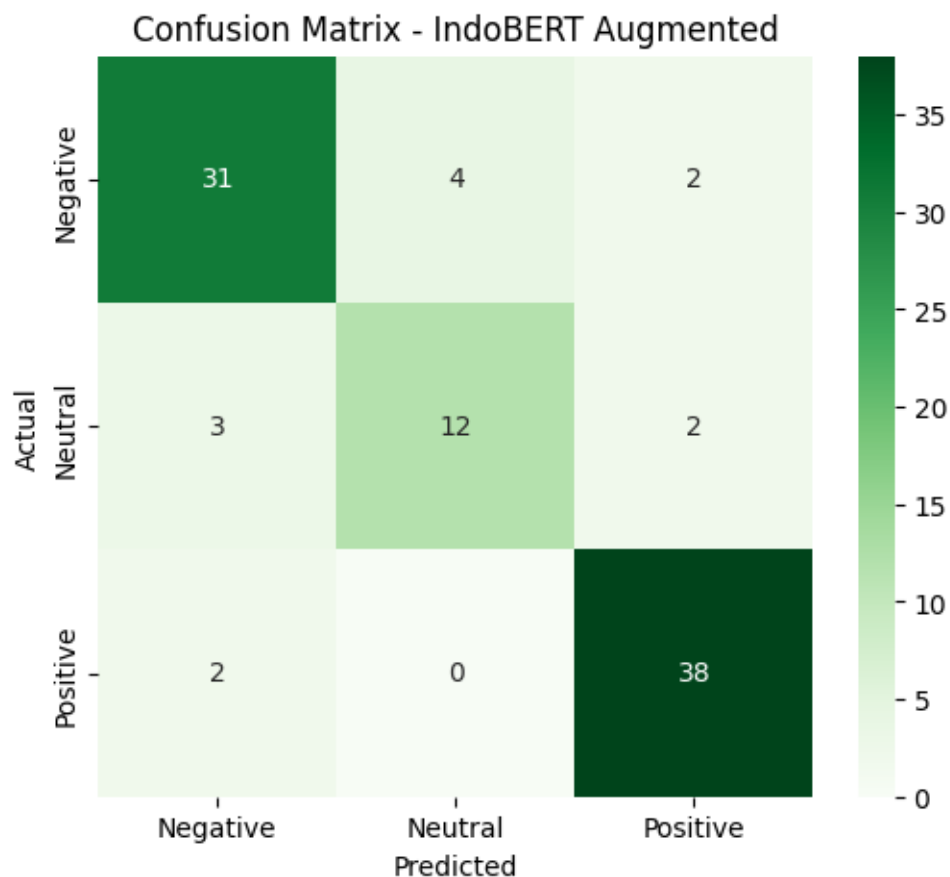
Counter({'Neutral': 161, 'Positive': 161, 'Negative': 161})

Distribusi Label Asli (y\_test):

Counter({'Positive': 40, 'Negative': 37, 'Neutral': 17})

Distribusi Prediksi (Augmented):

Counter({'Positive': 42, 'Negative': 36, 'Neutral': 16})



Metrik per Kelas (Augmented):

- Negative | Precision: 0.86 | Recall: 0.84 | F1: 0.85 | Support: 37.0
- Neutral | Precision: 0.75 | Recall: 0.71 | F1: 0.73 | Support: 17.0
- Positive | Precision: 0.90 | Recall: 0.95 | F1: 0.93 | Support: 40.0

Macro Avg F1-Score : 0.8345

Weighted Avg F1-Score: 0.8602

#6: Perbandingan Lanjutan BoW+NB

```
[103]: import pandas as pd

labels = ['Negative', 'Neutral', 'Positive']
kelas_all = labels + ['macro avg', 'weighted avg']

# === Precision Table ===
precision_rows = []
for label in kelas_all:
    precision_rows.append({
```

```

        "Kelas": label.title() if "avg" in label else label,
        "Baseline": f"{report_bow_nb_asli[label]['precision']:.2f}",
        "Augmented": f"{report_aug_bert[label]['precision']:.2f}"
    })
df_precision = pd.DataFrame(precision_rows)
print("\n=== PERBANDINGAN PRECISION ===")
print(df_precision.to_string(index=False))

# === Recall Table ===
recall_rows = []
for label in kelas_all:
    recall_rows.append({
        "Kelas": label.title() if "avg" in label else label,
        "Baseline": f"{report_nb[label]['recall']:.2f}",
        "Augmented": f"{report_aug_bert[label]['recall']:.2f}"
    })
df_recall = pd.DataFrame(recall_rows)
print("\n=== PERBANDINGAN RECALL ===")
print(df_recall.to_string(index=False))

# === F1-Score Table ===
f1_rows = []
for label in kelas_all:
    f1_rows.append({
        "Kelas": label.title() if "avg" in label else label,
        "Baseline": f"{report_nb[label]['f1-score']:.2f}",
        "Augmented": f"{report_aug_bert[label]['f1-score']:.2f}"
    })
df_f1 = pd.DataFrame(f1_rows)
print("\n=== PERBANDINGAN F1-SCORE ===")
print(df_f1.to_string(index=False))

# Buat tabel overall metrik
overall_rows = [
    {
        "Metrik": "Accuracy",
        "Baseline": f"{accuracy_score(y_test, y_pred_baseline):.4f}",
        "Augmented": f"{accuracy_score(y_test, y_pred_aug_bert):.4f}"
    },
    {
        "Metrik": "Macro Precision",
        "Baseline": f"{report_nb['macro avg']['precision']:.2f}",
        "Augmented": f"{report_aug_bert['macro avg']['precision']:.2f}"
    },
    {
        "Metrik": "Macro Recall",

```



```

        "Baseline": f"{report_nb['macro avg']['recall']:.2f}",
        "Augmented": f"{report_aug_bert['macro avg']['recall']:.2f}"
    },
    {
        "Metrik": "Macro F1",
        "Baseline": f"{report_nb['macro avg']['f1-score']:.2f}",
        "Augmented": f"{report_aug_bert['macro avg']['f1-score']:.2f}"
    },
    {
        "Metrik": "Weighted Precision",
        "Baseline": f"{report_nb['weighted avg']['precision']:.2f}",
        "Augmented": f"{report_aug_bert['weighted avg']['precision']:.2f}"
    },
    {
        "Metrik": "Weighted Recall",
        "Baseline": f"{report_nb['weighted avg']['recall']:.2f}",
        "Augmented": f"{report_aug_bert['weighted avg']['recall']:.2f}"
    },
    {
        "Metrik": "Weighted F1",
        "Baseline": f"{report_nb['weighted avg']['f1-score']:.2f}",
        "Augmented": f"{report_aug_bert['weighted avg']['f1-score']:.2f}"
    }
]

df_overall = pd.DataFrame(overall_rows)
print("\n=== TABEL OVERALL METRIK BASELINE vs AUGMENTED ===")
print(df_overall.to_string(index=False))

```

=== PERBANDINGAN PRECISION ===

Kelas	Baseline	Augmented
Negative	0.82	0.86
Neutral	0.75	0.75
Positive	0.88	0.90
Macro Avg	0.82	0.84
Weighted Avg	0.83	0.86

=== PERBANDINGAN RECALL ===

Kelas	Baseline	Augmented
Negative	0.86	0.84
Neutral	0.53	0.71
Positive	0.95	0.95
Macro Avg	0.78	0.83
Weighted Avg	0.84	0.86

=== PERBANDINGAN F1-SCORE ===

	Kelas	Baseline	Augmented
	Negative	0.84	0.85
	Neutral	0.62	0.73
	Positive	0.92	0.93
	Macro Avg	0.79	0.83
	Weighted Avg	0.83	0.86

=== TABEL OVERALL METRIK BASELINE vs AUGMENTED ===

	Metrik	Baseline	Augmented
	Accuracy	0.8404	0.8617
	Macro Precision	0.82	0.84
	Macro Recall	0.78	0.83
	Macro F1	0.79	0.83
	Weighted Precision	0.83	0.86
	Weighted Recall	0.84	0.86
	Weighted F1	0.83	0.86

#7: Eksplor BoW + LR Bag of Words, Linear Regression

```
[104]: from sklearn.pipeline import Pipeline
from sklearn.feature_extraction.text import CountVectorizer
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import classification_report, accuracy_score

# === BoW + LR (Asli) ===
pipeline_bow_lr = Pipeline([
    ('bow', CountVectorizer()),
    ('clf', LogisticRegression(max_iter=1000))
])
pipeline_bow_lr.fit(X_train, y_train)
y_pred_bow_lr = pipeline_bow_lr.predict(X_test)
report_bow_lr = classification_report(y_test, y_pred_bow_lr, output_dict=True,
    ↪zero_division=0)
acc_bow_lr = accuracy_score(y_test, y_pred_bow_lr)

# === BoW + LR (Augmented) ===
pipeline_bow_lr_aug = Pipeline([
    ('bow', CountVectorizer()),
    ('clf', LogisticRegression(max_iter=1000))
])
pipeline_bow_lr_aug.fit(X_train_aug_bert, y_train_aug_bert)
y_pred_bow_lr_aug = pipeline_bow_lr_aug.predict(X_test)
report_bow_lr_aug = classification_report(y_test, y_pred_bow_lr_aug,
    ↪output_dict=True, zero_division=0)
acc_bow_lr_aug = accuracy_score(y_test, y_pred_bow_lr_aug)

# === Tampilkan Perbandingan ===
print("\n=== BoW + Logistic Regression ===")
```

```

print(f"Akurasi (Asli)      : {acc_bow_lr:.4f}")
print(f"Akurasi (Augmented): {acc_bow_lr_aug:.4f}\n")

for label in ['Negative', 'Neutral', 'Positive']:
    print(f"{label} Class:")
    print(f"  F1 Asli      : {report_bow_lr[label]['f1-score']:.2f}")
    print(f"  F1 Augmented: {report_bow_lr_aug[label]['f1-score']:.2f}")

print(f"\nMacro F1 Asli      : {report_bow_lr['macro avg']['f1-score']:.4f}")
print(f"Macro F1 Augmented: {report_bow_lr_aug['macro avg']['f1-score']:.4f}")

```

=== BoW + Logistic Regression ===

Akurasi (Asli) : 0.7872

Akurasi (Augmented): 0.7872

Negative Class:

F1 Asli : 0.85

F1 Augmented: 0.85

Neutral Class:

F1 Asli : 0.43

F1 Augmented: 0.50

Positive Class:

F1 Asli : 0.85

F1 Augmented: 0.84

Macro F1 Asli : 0.7095

Macro F1 Augmented: 0.7307

#7: Eksplor BoW + SVM Bag of Words, Support Vector Machine

```

[105]: from sklearn.pipeline import Pipeline
from sklearn.feature_extraction.text import CountVectorizer
from sklearn.svm import LinearSVC
from sklearn.metrics import classification_report, accuracy_score

# === BoW + SVM (Asli) ===
pipeline_bow_svm = Pipeline([
    ('bow', CountVectorizer()),
    ('clf', LinearSVC())
])
pipeline_bow_svm.fit(X_train, y_train)
y_pred_bow_svm = pipeline_bow_svm.predict(X_test)
report_bow_svm = classification_report(y_test, y_pred_bow_svm,
    ↳output_dict=True, zero_division=0)
acc_bow_svm = accuracy_score(y_test, y_pred_bow_svm)

# === BoW + SVM (Augmented) ===

```

```

pipeline_bow_svm_aug = Pipeline([
    ('bow', CountVectorizer()),
    ('clf', LinearSVC())
])
pipeline_bow_svm_aug.fit(X_train_aug_bert, y_train_aug_bert)
y_pred_bow_svm_aug = pipeline_bow_svm_aug.predict(X_test)
report_bow_svm_aug = classification_report(y_test, y_pred_bow_svm_aug,
    ↪output_dict=True, zero_division=0)
acc_bow_svm_aug = accuracy_score(y_test, y_pred_bow_svm_aug)

# === Tampilkan Perbandingan ===
print("\n=== BoW + SVM ===")
print(f"Akurasi (Asli)      : {acc_bow_svm:.4f}")
print(f"Akurasi (Augmented): {acc_bow_svm_aug:.4f}\n")

for label in ['Negative', 'Neutral', 'Positive']:
    print(f"{label} Class:")
    print(f"  F1 Asli      : {report_bow_svm[label]['f1-score']:.2f}")
    print(f"  F1 Augmented: {report_bow_svm_aug[label]['f1-score']:.2f}")

print(f"\nMacro F1 Asli      : {report_bow_svm['macro avg']['f1-score']:.4f}")
print(f"Macro F1 Augmented: {report_bow_svm_aug['macro avg']['f1-score']:.4f}")

```

=== BoW + SVM ===

Akurasi (Asli) : 0.7766

Akurasi (Augmented): 0.7872

Negative Class:

F1 Asli : 0.84

F1 Augmented: 0.85

Neutral Class:

F1 Asli : 0.43

F1 Augmented: 0.48

Positive Class:

F1 Asli : 0.84

F1 Augmented: 0.84

Macro F1 Asli : 0.7012

Macro F1 Augmented: 0.7227

/usr/local/lib/python3.12/dist-packages/sklearn/svm/\_base.py:1249:

ConvergenceWarning: Liblinear failed to converge, increase the number of iterations.

warnings.warn(

#8: Eksplor TF-IDF + NB Bag of Words, Naive Bayes

```
[106]: from sklearn.pipeline import Pipeline
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.naive_bayes import MultinomialNB
from sklearn.metrics import classification_report, accuracy_score

# === TF-IDF + Naïve Bayes (Tanpa Augmentasi) ===
pipeline_nb = Pipeline([
    ('tfidf', TfidfVectorizer()),
    ('clf', MultinomialNB())
])
pipeline_nb.fit(X_train, y_train)
y_pred_nb = pipeline_nb.predict(X_test)
#report_nb = classification_report(y_test, y_pred_nb, output_dict=True,
    ↪zero_division=0)
#acc_nb = accuracy_score(y_test, y_pred_nb)
report_tfidf_nb_asli = classification_report(y_test, y_pred_nb,
    ↪output_dict=True, zero_division=0)
acc_tfidf_nb_asli = accuracy_score(y_test, y_pred_nb)

# === TF-IDF + Naïve Bayes (Dengan Augmentasi) ===
pipeline_nb_aug = Pipeline([
    ('tfidf', TfidfVectorizer()),
    ('clf', MultinomialNB())
])
pipeline_nb_aug.fit(X_train_aug_bert, y_train_aug_bert)
y_pred_nb_aug = pipeline_nb_aug.predict(X_test)
#report_nb_aug = classification_report(y_test, y_pred_nb_aug, output_dict=True,
    ↪zero_division=0)
#acc_nb_aug = accuracy_score(y_test, y_pred_nb_aug)
report_tfidf_nb_aug = classification_report(y_test, y_pred_nb_aug,
    ↪output_dict=True, zero_division=0)
acc_tfidf_nb_aug = accuracy_score(y_test, y_pred_nb_aug)

# === Tampilkan Perbandingan ===
print("\n=== TF-IDF + Naïve Bayes ===")
print(f"Akurasi (Asli)      : {acc_tfidf_nb_asli:.4f}")
print(f"Akurasi (Augmented): {acc_tfidf_nb_aug:.4f}\n")

for label in ['Negative', 'Neutral', 'Positive']:
    print(f"{label} Class:")
    print(f"  F1 Asli      : {report_nb[label]['f1-score']:.2f}")
    print(f"  F1 Augmented: {report_nb_aug[label]['f1-score']:.2f}")

print(f"\nMacro F1 Asli      : {report_nb['macro avg']['f1-score']:.4f}")
print(f"Macro F1 Augmented: {report_nb_aug['macro avg']['f1-score']:.4f}")
```

=== TF-IDF + Naïve Bayes ===

Akurasi (Asli) : 0.7447

Akurasi (Augmented): 0.8404

Negative Class:

F1 Asli : 0.84

F1 Augmented: 0.83

Neutral Class:

F1 Asli : 0.62

F1 Augmented: 0.67

Positive Class:

F1 Asli : 0.92

F1 Augmented: 0.93

Macro F1 Asli : 0.7928

Macro F1 Augmented: 0.8083

#9: Eksplor TF-IDF + LR Bag of Words, Linear Regression

```
[107]: from sklearn.linear_model import LogisticRegression

# === TF-IDF + Logistic Regression (Tanpa Augmentasi) ===
pipeline_lr = Pipeline([
    ('tfidf', TfidfVectorizer()),
    ('clf', LogisticRegression(max_iter=1000))
])
pipeline_lr.fit(X_train, y_train)
y_pred_lr = pipeline_lr.predict(X_test)
report_lr = classification_report(y_test, y_pred_lr, output_dict=True,
    ↪zero_division=0)
acc_lr = accuracy_score(y_test, y_pred_lr)

# === TF-IDF + Logistic Regression (Dengan Augmentasi) ===
pipeline_lr_aug = Pipeline([
    ('tfidf', TfidfVectorizer()),
    ('clf', LogisticRegression(max_iter=1000))
])
pipeline_lr_aug.fit(X_train_aug_bert, y_train_aug_bert)
y_pred_lr_aug = pipeline_lr_aug.predict(X_test)
report_lr_aug = classification_report(y_test, y_pred_lr_aug, output_dict=True,
    ↪zero_division=0)
acc_lr_aug = accuracy_score(y_test, y_pred_lr_aug)

# === Tampilkan Perbandingan ===
print("\n=== TF-IDF + Logistic Regression ===")
print(f"Akurasi (Asli) : {acc_lr:.4f}")
print(f"Akurasi (Augmented): {acc_lr_aug:.4f}\n")
```

```

for label in ['Negative', 'Neutral', 'Positive']:
    print(f"{label} Class:")
    print(f"    F1 Asli      : {report_lr[label]['f1-score']:.2f}")
    print(f"    F1 Augmented: {report_lr_aug[label]['f1-score']:.2f}")

print(f"\nMacro F1 Asli      : {report_lr['macro avg']['f1-score']:.4f}")
print(f"Macro F1 Augmented: {report_lr_aug['macro avg']['f1-score']:.4f}")

```

=== TF-IDF + Logistic Regression ===

Akurasi (Asli) : 0.7340

Akurasi (Augmented): 0.8085

Negative Class:

F1 Asli : 0.80

F1 Augmented: 0.82

Neutral Class:

F1 Asli : 0.00

F1 Augmented: 0.50

Positive Class:

F1 Asli : 0.81

F1 Augmented: 0.88

Macro F1 Asli : 0.5377

Macro F1 Augmented: 0.7353

#10: Eksplor TF-IDF + SVM Bag of Words, Support Vector Machine

```

[108]: from sklearn.svm import LinearSVC

# === TF-IDF + SVM (Tanpa Augmentasi) ===
pipeline_svm = Pipeline([
    ('tfidf', TfidfVectorizer()),
    ('clf', LinearSVC())
])
pipeline_svm.fit(X_train, y_train)
y_pred_svm = pipeline_svm.predict(X_test)
report_svm = classification_report(y_test, y_pred_svm, output_dict=True,
    ↪zero_division=0)
acc_svm = accuracy_score(y_test, y_pred_svm)

# === TF-IDF + SVM (Dengan Augmentasi) ===
pipeline_svm_aug = Pipeline([
    ('tfidf', TfidfVectorizer()),
    ('clf', LinearSVC())
])
pipeline_svm_aug.fit(X_train_aug_bert, y_train_aug_bert)

```

```

y_pred_svm_aug = pipeline_svm_aug.predict(X_test)
report_svm_aug = classification_report(y_test, y_pred_svm_aug,
    ↳output_dict=True, zero_division=0)
acc_svm_aug = accuracy_score(y_test, y_pred_svm_aug)

# === Tampilkan Perbandingan ===
print("\n=== TF-IDF + SVM ===")
print(f"Akurasi (Asli)      : {acc_svm:.4f}")
print(f"Akurasi (Augmented): {acc_svm_aug:.4f}\n")

for label in ['Negative', 'Neutral', 'Positive']:
    print(f"{label} Class:")
    print(f"  F1 Asli      : {report_svm[label]['f1-score']:.2f}")
    print(f"  F1 Augmented: {report_svm_aug[label]['f1-score']:.2f}")

print(f"\nMacro F1 Asli      : {report_svm['macro avg']['f1-score']:.4f}")
print(f"Macro F1 Augmented: {report_svm_aug['macro avg']['f1-score']:.4f}")

```

```

=== TF-IDF + SVM ===
Akurasi (Asli)      : 0.7979
Akurasi (Augmented): 0.8191

```

```

Negative Class:
  F1 Asli      : 0.84
  F1 Augmented: 0.84
Neutral Class:
  F1 Asli      : 0.36
  F1 Augmented: 0.50
Positive Class:
  F1 Asli      : 0.87
  F1 Augmented: 0.89

```

```

Macro F1 Asli      : 0.6912
Macro F1 Augmented: 0.7437

```

#11: Final All-in-one

## 1.1 Confusion Matrix

```

[109]: import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.metrics import confusion_matrix
from sklearn.pipeline import Pipeline
from sklearn.feature_extraction.text import CountVectorizer, TfidfVectorizer
from sklearn.naive_bayes import MultinomialNB
from sklearn.linear_model import LogisticRegression
from sklearn.svm import LinearSVC

```



```

pipeline_bow_nb_aug = Pipeline([
    ('vect', CountVectorizer()),
    ('clf', MultinomialNB())
])
pipeline_bow_nb_aug.fit(X_train_aug_bert, y_train_aug_bert)

pipeline_bow_lr_aug = Pipeline([
    ('bow', CountVectorizer()),
    ('clf', LogisticRegression(max_iter=1000))
])
pipeline_bow_lr_aug.fit(X_train_aug_bert, y_train_aug_bert)

pipeline_bow_svm_aug = Pipeline([
    ('bow', CountVectorizer()),
    ('clf', LinearSVC())
])
pipeline_bow_svm_aug.fit(X_train_aug_bert, y_train_aug_bert)

# Daftar model dan prediksi
model_preds = {
    "BoW + NB (Asli)": pipeline_bow_nb_asli.predict(X_test),
    "BoW + NB (Augmented)": pipeline_bow_nb_aug.predict(X_test),
    "BoW + LR (Asli)": pipeline_bow_lr.predict(X_test),
    "BoW + LR (Augmented)": pipeline_bow_lr_aug.predict(X_test),
    "BoW + SVM (Asli)": pipeline_bow_svm.predict(X_test),
    "BoW + SVM (Augmented)": pipeline_bow_svm_aug.predict(X_test),
    "TF-IDF + NB (Asli)": pipeline_tfidf_nb.predict(X_test),
    "TF-IDF + NB (Augmented)": pipeline_tfidf_nb_aug.predict(X_test),
    "TF-IDF + LR (Asli)": pipeline_tfidf_lr.predict(X_test),
    "TF-IDF + LR (Augmented)": pipeline_tfidf_lr_aug.predict(X_test),
    "TF-IDF + SVM (Asli)": pipeline_tfidf_svm.predict(X_test),
    "TF-IDF + SVM (Augmented)": pipeline_tfidf_svm_aug.predict(X_test),
}

labels = ["Negative", "Neutral", "Positive"]

# Plot grid
fig, axes = plt.subplots(nrows=4, ncols=3, figsize=(18, 20))
fig.suptitle("Confusion Matrix Semua Model", fontsize=18)

for ax, (model_name, y_pred) in zip(axes.flat, model_preds.items()):
    cm = confusion_matrix(y_test, y_pred, labels=labels)
    sns.heatmap(cm, annot=True, fmt='d', cmap='Blues', xticklabels=labels,
    yticklabels=labels, ax=ax)
    ax.set_title(model_name)
    ax.set_xlabel("Predicted")

```

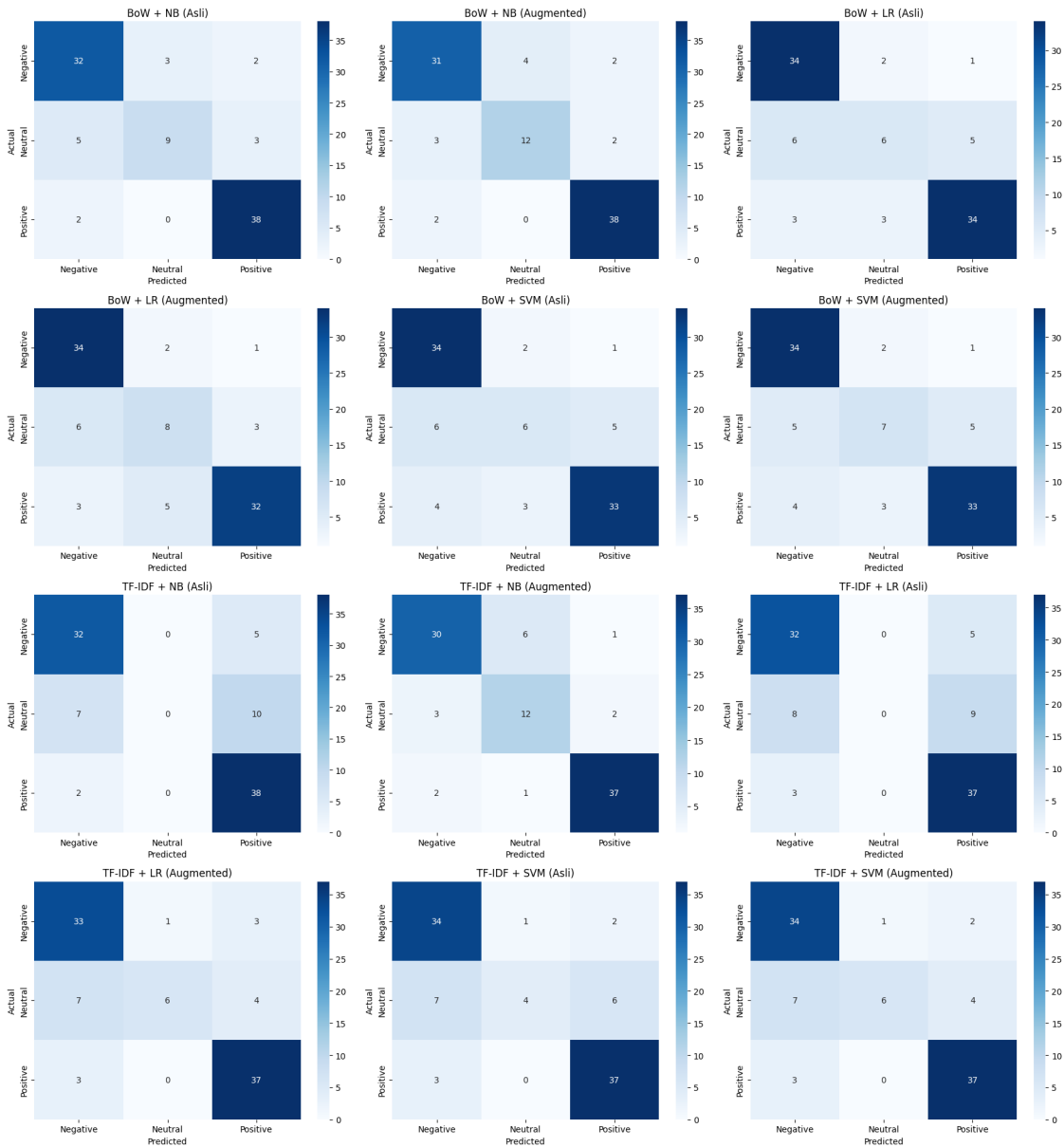
```
ax.set_ylabel("Actual")
```

```
plt.tight_layout(rect=[0, 0, 1, 0.97])
plt.show()
```

/usr/local/lib/python3.12/dist-packages/sklearn/svm/\_base.py:1249:  
ConvergenceWarning: Liblinear failed to converge, increase the number of  
iterations.

```
warnings.warn(
```

Confusion Matrix Semua Model



## 1.2 Tabel Perbandingan

```
[111]: import pandas as pd
from sklearn.metrics import classification_report, accuracy_score

# Simpan semua report dan akurasi
report_bow_nb = classification_report(y_test, pipeline_bow_nb_asli.
    ↳predict(X_test), output_dict=True, zero_division=0)
acc_bow_nb = accuracy_score(y_test, pipeline_bow_nb_asli.predict(X_test))

report_bow_nb_aug = classification_report(y_test, pipeline_bow_nb_aug.
    ↳predict(X_test), output_dict=True, zero_division=0)
acc_bow_nb_aug = accuracy_score(y_test, pipeline_bow_nb_aug.predict(X_test))

report_bow_lr = classification_report(y_test, pipeline_bow_lr.predict(X_test),
    ↳output_dict=True, zero_division=0)
acc_bow_lr = accuracy_score(y_test, pipeline_bow_lr.predict(X_test))

report_bow_lr_aug = classification_report(y_test, pipeline_bow_lr_aug.
    ↳predict(X_test), output_dict=True, zero_division=0)
acc_bow_lr_aug = accuracy_score(y_test, pipeline_bow_lr_aug.predict(X_test))

report_bow_svm = classification_report(y_test, pipeline_bow_svm.
    ↳predict(X_test), output_dict=True, zero_division=0)
acc_bow_svm = accuracy_score(y_test, pipeline_bow_svm.predict(X_test))

report_bow_svm_aug = classification_report(y_test, pipeline_bow_svm_aug.
    ↳predict(X_test), output_dict=True, zero_division=0)
acc_bow_svm_aug = accuracy_score(y_test, pipeline_bow_svm_aug.predict(X_test))

report_tfidf_nb = classification_report(y_test, pipeline_tfidf_nb.
    ↳predict(X_test), output_dict=True, zero_division=0)
acc_tfidf_nb = accuracy_score(y_test, pipeline_tfidf_nb.predict(X_test))

report_tfidf_nb_aug = classification_report(y_test, pipeline_tfidf_nb_aug.
    ↳predict(X_test), output_dict=True, zero_division=0)
acc_tfidf_nb_aug = accuracy_score(y_test, pipeline_tfidf_nb_aug.predict(X_test))

report_tfidf_lr = classification_report(y_test, pipeline_tfidf_lr.
    ↳predict(X_test), output_dict=True, zero_division=0)
acc_tfidf_lr = accuracy_score(y_test, pipeline_tfidf_lr.predict(X_test))

report_tfidf_lr_aug = classification_report(y_test, pipeline_tfidf_lr_aug.
    ↳predict(X_test), output_dict=True, zero_division=0)
acc_tfidf_lr_aug = accuracy_score(y_test, pipeline_tfidf_lr_aug.predict(X_test))
```

```

report_tfidf_svm = classification_report(y_test, pipeline_tfidf_svm.
    ↪predict(X_test), output_dict=True, zero_division=0)
acc_tfidf_svm = accuracy_score(y_test, pipeline_tfidf_svm.predict(X_test))

report_tfidf_svm_aug = classification_report(y_test, pipeline_tfidf_svm_aug.
    ↪predict(X_test), output_dict=True, zero_division=0)
acc_tfidf_svm_aug = accuracy_score(y_test, pipeline_tfidf_svm_aug.
    ↪predict(X_test))

# Helper function
def extract_metrics(report, acc):
    return {
        "Accuracy": acc,
        "Macro F1": report["macro avg"]["f1-score"],
        "Negative Precision": report["Negative"]["precision"],
        "Negative Recall": report["Negative"]["recall"],
        "Negative F1": report["Negative"]["f1-score"],
        "Negative Support": report["Negative"]["support"],
        "Neutral Precision": report["Neutral"]["precision"],
        "Neutral Recall": report["Neutral"]["recall"],
        "Neutral F1": report["Neutral"]["f1-score"],
        "Neutral Support": report["Neutral"]["support"],
        "Positive Precision": report["Positive"]["precision"],
        "Positive Recall": report["Positive"]["recall"],
        "Positive F1": report["Positive"]["f1-score"],
        "Positive Support": report["Positive"]["support"]
    }

# Semua model
models = [
    ("BoW", "Naïve Bayes", "Asli", extract_metrics(report_bow_nb, acc_bow_nb)),
    ("BoW", "Naïve Bayes", "Augmented", extract_metrics(report_bow_nb_aug,
    ↪acc_bow_nb_aug)),
    ("BoW", "Logistic Reg.", "Asli", extract_metrics(report_bow_lr,
    ↪acc_bow_lr)),
    ("BoW", "Logistic Reg.", "Augmented", extract_metrics(report_bow_lr_aug,
    ↪acc_bow_lr_aug)),
    ("BoW", "SVM", "Asli", extract_metrics(report_bow_svm, acc_bow_svm)),
    ("BoW", "SVM", "Augmented", extract_metrics(report_bow_svm_aug,
    ↪acc_bow_svm_aug)),
    ("TF-IDF", "Naïve Bayes", "Asli", extract_metrics(report_tfidf_nb,
    ↪acc_tfidf_nb)),
    ("TF-IDF", "Naïve Bayes", "Augmented", extract_metrics(report_tfidf_nb_aug,
    ↪acc_tfidf_nb_aug)),
    ("TF-IDF", "Logistic Reg.", "Asli", extract_metrics(report_tfidf_lr,
    ↪acc_tfidf_lr)),

```

```

        ("TF-IDF", "Logistic Reg.", "Augmented",
        ↪extract_metrics(report_tfidf_lr_aug, acc_tfidf_lr_aug)),
        ("TF-IDF", "SVM", "Asli", extract_metrics(report_tfidf_svm, acc_tfidf_svm)),
        ("TF-IDF", "SVM", "Augmented", extract_metrics(report_tfidf_svm_aug,
        ↪acc_tfidf_svm_aug)),
    ]

# Tabel per kelas
df_negative = pd.DataFrame([
    [f, a, d,
     round(m["Negative Precision"], 2),
     round(m["Negative Recall"], 2),
     round(m["Negative F1"], 2),
     int(m["Negative Support"])]
    for f, a, d, m in models
], columns=["Fitur", "Algoritma", "Data", "Precision", "Recall", "F1 Negative",
↪"Support"])

df_neutral = pd.DataFrame([
    [f, a, d,
     round(m["Neutral Precision"], 2),
     round(m["Neutral Recall"], 2),
     round(m["Neutral F1"], 2),
     int(m["Neutral Support"])]
    for f, a, d, m in models
], columns=["Fitur", "Algoritma", "Data", "Precision", "Recall", "F1 Neutral",
↪"Support"])

df_positive = pd.DataFrame([
    [f, a, d,
     round(m["Positive Precision"], 2),
     round(m["Positive Recall"], 2),
     round(m["Positive F1"], 2),
     int(m["Positive Support"])]
    for f, a, d, m in models
], columns=["Fitur", "Algoritma", "Data", "Precision", "Recall", "F1 Positive",
↪"Support"])

df_overall = pd.DataFrame([
    [f, a, d, round(m["Accuracy"], 4), round(m["Macro F1"], 4)]
    for f, a, d, m in models
], columns=["Fitur", "Algoritma", "Data", "Accuracy", "Macro F1"])

# Tampilkan semua
print("\n=== TABEL NEGATIVE ===")
print(df_negative.to_string(index=False))

```

```

print("\n=== TABEL NEUTRAL ===")
print(df_neutral.to_string(index=False))

print("\n=== TABEL POSITIVE ===")
print(df_positive.to_string(index=False))

print("\n=== TABEL OVERALL METRIK ===")
print(df_overall.to_string(index=False))

```

=== TABEL NEGATIVE ===

Fitur	Algoritma	Data	Precision	Recall	F1 Negative	Support
BoW	Naïve Bayes	Asli	0.82	0.86	0.84	37
BoW	Naïve Bayes	Augmented	0.86	0.84	0.85	37
BoW	Logistic Reg.	Asli	0.79	0.92	0.85	37
BoW	Logistic Reg.	Augmented	0.79	0.92	0.85	37
BoW	SVM	Asli	0.77	0.92	0.84	37
BoW	SVM	Augmented	0.79	0.92	0.85	37
TF-IDF	Naïve Bayes	Asli	0.78	0.86	0.82	37
TF-IDF	Naïve Bayes	Augmented	0.86	0.81	0.83	37
TF-IDF	Logistic Reg.	Asli	0.74	0.86	0.80	37
TF-IDF	Logistic Reg.	Augmented	0.77	0.89	0.82	37
TF-IDF	SVM	Asli	0.77	0.92	0.84	37
TF-IDF	SVM	Augmented	0.77	0.92	0.84	37

=== TABEL NEUTRAL ===

Fitur	Algoritma	Data	Precision	Recall	F1 Neutral	Support
BoW	Naïve Bayes	Asli	0.75	0.53	0.62	17
BoW	Naïve Bayes	Augmented	0.75	0.71	0.73	17
BoW	Logistic Reg.	Asli	0.55	0.35	0.43	17
BoW	Logistic Reg.	Augmented	0.53	0.47	0.50	17
BoW	SVM	Asli	0.55	0.35	0.43	17
BoW	SVM	Augmented	0.58	0.41	0.48	17
TF-IDF	Naïve Bayes	Asli	0.00	0.00	0.00	17
TF-IDF	Naïve Bayes	Augmented	0.63	0.71	0.67	17
TF-IDF	Logistic Reg.	Asli	0.00	0.00	0.00	17
TF-IDF	Logistic Reg.	Augmented	0.86	0.35	0.50	17
TF-IDF	SVM	Asli	0.80	0.24	0.36	17
TF-IDF	SVM	Augmented	0.86	0.35	0.50	17

=== TABEL POSITIVE ===

Fitur	Algoritma	Data	Precision	Recall	F1 Positive	Support
BoW	Naïve Bayes	Asli	0.88	0.95	0.92	40
BoW	Naïve Bayes	Augmented	0.90	0.95	0.93	40
BoW	Logistic Reg.	Asli	0.85	0.85	0.85	40
BoW	Logistic Reg.	Augmented	0.89	0.80	0.84	40
BoW	SVM	Asli	0.85	0.82	0.84	40

BoW	SVM Augmented	0.85	0.82	0.84	40
TF-IDF	Naïve Bayes Asli	0.72	0.95	0.82	40
TF-IDF	Naïve Bayes Augmented	0.93	0.93	0.93	40
TF-IDF	Logistic Reg. Asli	0.73	0.93	0.81	40
TF-IDF	Logistic Reg. Augmented	0.84	0.93	0.88	40
TF-IDF	SVM Asli	0.82	0.93	0.87	40
TF-IDF	SVM Augmented	0.86	0.93	0.89	40

=== TABEL OVERALL METRIK ===

Fitur	Algoritma	Data	Accuracy	Macro F1
BoW	Naïve Bayes	Asli	0.8404	0.7928
BoW	Naïve Bayes	Augmented	0.8617	0.8345
BoW	Logistic Reg.	Asli	0.7872	0.7095
BoW	Logistic Reg.	Augmented	0.7872	0.7307
BoW	SVM	Asli	0.7766	0.7012
BoW	SVM	Augmented	0.7872	0.7227
TF-IDF	Naïve Bayes	Asli	0.7447	0.5459
TF-IDF	Naïve Bayes	Augmented	0.8404	0.8083
TF-IDF	Logistic Reg.	Asli	0.7340	0.5377
TF-IDF	Logistic Reg.	Augmented	0.8085	0.7353
TF-IDF	SVM	Asli	0.7979	0.6912
TF-IDF	SVM	Augmented	0.8191	0.7437

## 2 CSV Mentahan

### 2.1 Augmented + Train/test

```
[113]: import pandas as pd

# Gabungkan data latih dan uji
df_train = pd.DataFrame({
    "text": X_train_aug_bert,
    "label": y_train_aug_bert,
    "split": "train"
})

df_test = pd.DataFrame({
    "text": X_test,
    "label": y_test,
    "split": "test"
})

# Gabung jadi satu DataFrame
df_combined = pd.concat([df_train, df_test], ignore_index=True)

# Simpan ke CSV
df_combined.to_csv("garuda_combined_split.csv", index=False)
```

```
print(f" CSV gabungan disimpan sebagai 'garuda_combined_split.csv'␣
↪({len(df_combined)} baris)")
```

CSV gabungan disimpan sebagai 'garuda\_combined\_split.csv' (577 baris)

## 2.2 + Label augmented/original

```
[114]: import pandas as pd

# Data latih asli
df_original = pd.DataFrame({
    "text": X_train.reset_index(drop=True),
    "label": y_train.reset_index(drop=True),
    "source": "original"
})

# Data hasil augmentasi
df_augmented = pd.DataFrame({
    "text": pd.Series(new_neutral_texts + new_negative_texts),
    "label": pd.Series(['Neutral'] * len(new_neutral_texts) + ['Negative'] *␣
↪len(new_negative_texts)),
    "source": "augmented"
})

# Gabungkan
df_augmented_flagged = pd.concat([df_original, df_augmented], ignore_index=True)

# Simpan ke CSV
df_augmented_flagged.to_csv("garuda_augmented_flagged.csv", index=False)
print(f" CSV gabungan disimpan sebagai 'garuda_augmented_flagged.csv'␣
↪({len(df_augmented_flagged)} baris)")
```

CSV gabungan disimpan sebagai 'garuda\_augmented\_flagged.csv' (483 baris)