

Penerapan Metode Naïve Bayes Classifier terhadap Sentimen Masyarakat pada Tagar #Kaburajadulu di Media Sosial Twitter

Eni Purnama, Bayu Suriaatmaja Suwanda

^{1,2}Komunikasi Digital dan Media, Institut Pertanian Bogor, Jl. Kumbang No.14, RT.02/RW/06, Babakan, Kota Bogor, 16128

¹E-mail : eni09purnama@apps.ipb.ac.id

²Email : bayu.suwanda@apps.ipb.ac.id

Abstract— Twitter is a social media platform that allows its users to communicate in short messages and has become a space for expressing public opinion, including through the popular hashtag #kaburajadulu, which reflects the concerns of the young generation about social conditions and their dreams of seeking a better life abroad. This study aims to analyze public sentiment towards the hashtag using two approaches. First, sentiment classification was performed using the Naïve Bayes Classifier algorithm to divide tweets into positive, neutral, and negative categories. The process involved data collection, preprocessing, labeling, word weighting, classification, evaluation, and visualization. The classification results showed an accuracy of 56%. Second, a multinomial logistic regression analysis was conducted to determine the effect of three independent variables on sentiment. The results showed that all three variables had a significant influence, with varying degrees of contribution. These findings suggest that a combined approach between machine learning and statistical analysis can provide in-depth insights into communication patterns and public opinion on social issues emerging on social media, particularly in the context of migration and youth anxiety.

Keywords: *Sentiment Analysis; Communication; Naïve Bayes Classifier; Multinomial Logistic Regression*

Abstrak— Twitter merupakan media sosial yang memungkinkan penggunanya melakukan komunikasi dalam pesan singkat dan telah menjadi ruang ekspresi opini publik, termasuk melalui tagar populer #kaburajadulu yang mencerminkan keresahan generasi muda akan kondisi sosial dan impian mencari kehidupan lebih baik di luar negeri. Penelitian ini bertujuan menganalisis sentimen masyarakat terhadap tagar tersebut dengan dua pendekatan. Pertama, dilakukan klasifikasi sentimen menggunakan algoritma Naïve Bayes Classifier untuk membagi tweet ke dalam kategori positif, netral, dan negatif. Prosesnya melibatkan tahapan pengumpulan data, preprocessing, pelabelan, pembobotan kata, klasifikasi, evaluasi, hingga visualisasi. Hasil klasifikasi menunjukkan akurasi sebesar 56%. Kedua, dilakukan analisis menggunakan Regresi Logistik Multinomial guna mengetahui pengaruh tiga variabel bebas terhadap sentimen. Hasilnya, ketiga variabel tersebut terbukti berpengaruh signifikan dengan kontribusi yang bervariasi. Temuan ini menunjukkan bahwa pendekatan kombinatorik antara machine learning dan analisis statistik dapat memberikan wawasan mendalam terhadap pola komunikasi serta opini publik mengenai isu sosial yang berkembang di media sosial, khususnya dalam konteks migrasi dan keresahan generasi muda.

Kata kunci : *Analisis Sentimen; Komunikasi; Naïve Bayes Classifier; Regresi Logistik Multinomial*

I. PENDAHULUAN

Perkembangan teknologi informasi telah memberikan pengaruh signifikan dalam dunia informasi dan komunikasi di era digital. Sejalan dengan perkembangan informasi dan komunikasi akses internet semakin mudah. Di Indonesia, sebanyak 221 juta jiwa penduduknya menggunakan internet dengan penggunaan untuk mengakses sosial media sebesar 79,5% (APJII, 2024). Media sosial merupakan produk yang dihasilkan dari sebuah media baru berbasis online melalui situs web yang dipergunakan untuk berbagi jaringan komunikasi seperti berbicara dan berpartisipasi (Hamidati dalam Rakhman *et al.* 2021).

Twitter adalah media sosial yang memungkinkan penggunanya melakukan komunikasi dalam pesan singkat dengan jumlah karakter maksimal 280 karakter (Putri *et al.* 2023). Pengguna Twitter biasanya menumpahkan apa yang mereka rasakan kedalam sebuah pesan

singkat. Pesan singkat dapat dikategorikan sebagai bentuk komunikasi interpersonal, terutama ketika seseorang membalas pesan singkat orang lain. Komunikasi interpersonal dilakukan untuk menyampaikan pesan atau informasi serta mendapatkan umpan balik secara langsung (Abdullah dan Ramadhan 2024). Komunikasi massa juga dapat ditemukan, yaitu ketika seseorang menyiarkan pesan yang diunggah. Komunikasi massa menurut Bungin dalam Hanum AJ (2025) adalah proses komunikasi yang dilakukan melalui media massa dengan berbagai tujuan komunikasi dan untuk menyampaikan informasi kepada khalayak luas. Selain menggunakan konsep komunikasi interpersonal dan komunikasi massa, penelitian ini berlandaskan pada *Action Assembly Theory* (AAT) yang dikemukakan oleh John O. Greene (2013) dalam bukunya *Message Production: Advances In Communication Theory*. Teori ini memberikan pandangan mendalam tentang bagaimana pesan komunikasi disusun dan disampaikan. Inti dari teori ini adalah dua konsep penting, yaitu "unit action" dan "action assembly," yang menjelaskan proses bagaimana pesan-pesan dikembangkan dan dipersembahkan dalam interaksi komunikasi (Fachlevi *et al.* 2024).

Baru-baru ini tagar #kaburajadulu menjadi populer di masyarakat setelah trending pada Januari 2025. Tagar ini mengemukakan opini mengenai kesenjangan sosial di Indonesia. Ungkapan kekecewaan sekaligus upaya masyarakat terutama anak muda untuk mendapatkan kesejahteraan hidup lebih layak dengan mencari pekerjaan di negara lain. Pernyataan itu menyebabkan munculnya berbagai komentar dan membuat masyarakat mulai berani berpendapat. Hal ini menyebabkan pro-kontra mengenai mendapatkan kesejahteraan hidup lebih layak dengan mencari pekerjaan di negara lain.

Dampak tagar ini, baik dalam hal peningkatan atau penurunan jumlah penempatan akan bergantung pada komunikasi yang terbentuk di media sosial Twitter dan sentimen masyarakat. Oleh karena itu, perlu dilakukan analisis lebih lanjut terhadap sentimen yang ditimbulkan oleh tagar ini di kalangan masyarakat. Agar hasil yang diinginkan jelas, maka dilakukan proses klasifikasi yang merujuk pada proses analisis sentimen. Analisis sentimen dinilai dari komentar publik terhadap postingan dari pengguna di media sosial (Satria dalam Lestari dan Nurisusilawati 2024). Penelitian ini mengadopsi metode Naïve Bayes Classifier dalam menganalisis sentimen. *Naïve Bayes Classifier* merupakan suatu pendekatan klasifikasi yang berdasarkan probabilitas kemunculan kata-kata dari suatu teks. Probabilitas dihitung dengan mengakumulasi frekuensi dan kombinasi nilai dari dataset yang digunakan.

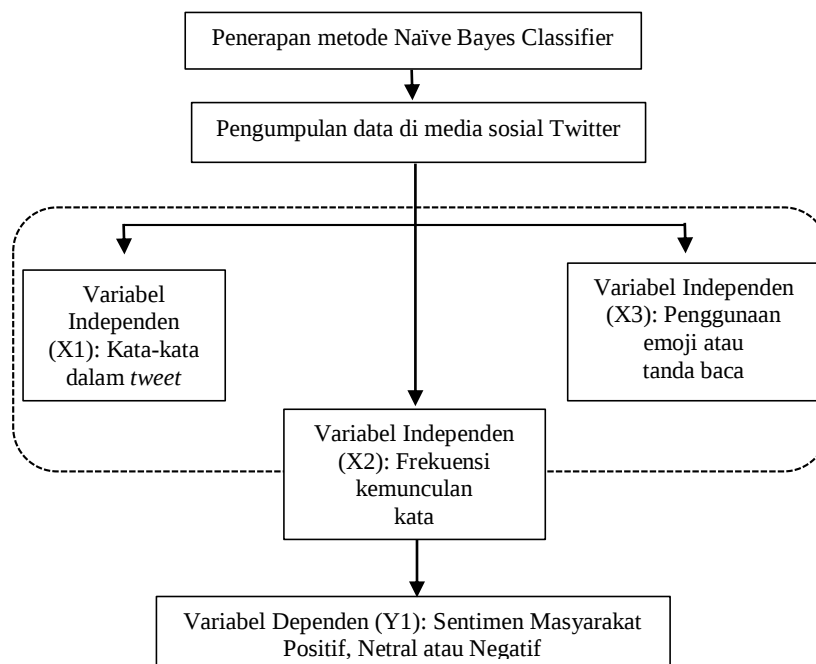
Penerapan metode *Naïve Bayes Classifier* telah diterapkan pada penelitian terdahulu oleh Agustin *et al.* (2024) yang melakukan penelitian dengan judul "Analisis Sentimen Menggunakan Metode *Naïve Bayes* Terhadap *Childfree*". Penelitian tersebut melihat bagaimana kelompok masyarakat dalam bentuk individual maupun kelompok menyebarkan sikap setuju maupun sikap tidak setuju. Kegiatan ini mengklasifikasikan sentimen masyarakat menjadi dua jenis klasifikasi, yaitu sentimen positif dan negatif. Metode analisis data yang digunakan adalah teknik klasifikasi dengan algoritma *Naïve Bayes Classifier* dengan hasil tingkat akurasi sebesar 96%. Penelitian lain yang dilakukan oleh Putri *et al.* (2023) dengan judul "Analisis Sentimen Terhadap Brand Skincare Lokal Menggunakan *Naïve Bayes Classifier*" menghasilkan nilai akurasi sebesar 79% untuk brand Avoskin, 78% untuk brand Azarine dan 75% untuk brand Somethinc. Atas pertimbangan dari dua penelitian tersebut, peneliti menyimpulkan metode yang dapat digunakan untuk mengatasi permasalahan di penelitian ini adalah *Naïve Bayes Classifier* yang bekerja berdasarkan probabilitas kemunculan kata-kata dari suatu teks. Namun, penelitian tersebut hanya fokus pada permasalahan *childfree* dan *brand skincare* lokal, sehingga belum ada penelitian yang membahas analisis sentimen masyarakat pada permasalahan yang diangkat dalam penelitian ini. Penelitian yang dilakukan oleh ZA Kusuma. (2024) dengan judul "Analisis Sentimen Pengguna Twitter Terhadap Program Makan Siang Gratis Menggunakan Metode *Naïve Bayes*" menghasilkan nilai akurasi

89%, presisi 90%, recall 90% dan F1-Score 89% dengan pembagian 80% untuk pelatihan dan 20% pengujian. Accuracy 89% menunjukkan bahwa keseluruhan model bekerja dengan baik dalam mengklasifikasikan tweet. Penelitian-penelitian terdahulu menjadi dasar pembaruan topik pada penelitian ini. Oleh karena itu, penelitian ini dilakukan untuk mengisi celah tersebut. Perbedaan yang dilakukan penulis dan peneliti terdahulu terletak pada pengembangan isu dan hasil yang diperoleh. Selain itu, guna memastikan keakuratan dan validitas data, penelitian ini menggunakan kuesioner sebagai instrumen pendukung dalam pengumpulan data yang akan disebarkan kepada responden pengguna media sosial Twitter.

Dengan demikian, penelitian ini memiliki urgensi untuk menganalisis penerapan metode *Naïve Bayes Classifier* terhadap sentimen masyarakat pada tagar #kaburajadulu di Media Sosial Twitter. Melalui hasil penelitian ini, diharapkan dapat lebih memahami bagaimana fenomena sosial tagar #kaburajadulu dilihat dari sudut pandang masyarakat, bagaimana media sosial Twitter menjadi saluran komunikasi, apa saja bentuk komunikasi dalam menyuarakan opini dan menilai performa metode *Naïve Bayes Classifier* dalam klasifikasi sentimen masyarakat.

II. METODE PELAKSANAAN

Kegiatan ini dilakukan di media sosial Twitter dalam rentang waktu 01 Februari 2025 hingga 31 Maret 2025. Berdasarkan pada latar belakang, kegiatan ini menggunakan dua pendekatan analisis. Pertama, dilakukan klasifikasi sentimen menggunakan algoritma *Naïve Bayes Classifier* untuk mengelompokkan tweet ke dalam tiga kategori: positif, netral, dan negatif. Proses dilakukan melalui tahapan pengumpulan data, *preprocessing*, *labeling*, pembobotan kata, klasifikasi, evaluasi dan visualisasi. Kedua, dilakukan analisis regresi logistik multinomial untuk melihat pengaruh variabel bebas terhadap sentimen. Oleh karena itu, terbentuk kerangka pikir yang merupakan jalur pemikiran yang dirancang berdasarkan kegiatan peneliti yang dilakukan. Kerangka berpikir merupakan model konseptual tentang bagaimana teori berhubungan dengan berbagai faktor yang telah diidentifikasi sebagai masalah yang penting (Sugiyono dalam Nurtikasari *et al.* 2022).



Gambar 1: Bagan konsep kerja

Maka dari itu hipotesis yang diajukan oleh penulis adalah sebagaimana berikut :
 H_1 : Terdapat pengaruh kata-kata dalam *tweet* (X1) terhadap (Y1).

H₂ : Terdapat frekuensi kemunculan kata tertentu (X₂) terhadap (Y₁).

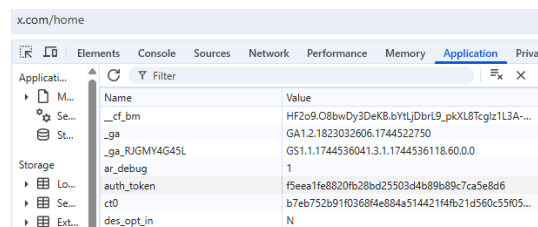
H₃ : Terdapat penggunaan emoji atau tanda baca (X₃) terhadap (Y₁).

III. HASIL DAN PEMBAHASAN

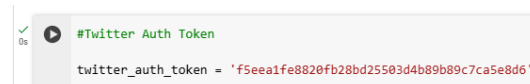
Naïve Bayes Classifier

Pengumpulan data

Proses pengumpulan data dengan teknik *crawling* data menggunakan *Python* serta *Library Python* yang dijalankan dengan *Google Colab*. Tahap pertama dalam pengumpulan data tweet dengan menghubungkan akun Twitter pribadi dengan *Google Colab* kemudian mencantumkan kode auth-token akun Twitter pribadi ke dalam *notebook Google Colab*.



Gambar 2. Kode auth-token akun Twitter pribadi



Gambar 3. Kode auth-token dihubungkan ke dalam *Google Colab*

Pengumpulan data dilakukan terhadap tweet berbahasa Indonesia dengan *keyword* “#kaburajadulu”. Proses pengambilan data dilakukan dari tanggal 1 Februari 2025 sampai 31 Maret 2025 dan mendapatkan sebanyak 1000 dataset dengan format csv.

Preprocessing

Pada tahap *preprocessing*, data dibersihkan dari elemen-elemen yang tidak relevan seperti simbol, angka, tanda baca, kata tidak baku, dan kesalahan pengetikan. Tujuan dilakukan *preprocessing* adalah memudahkan dalam pengolahan data dan membuat hasil data memiliki akurasi yang baik. Tahap *preprocessing* terdiri dari beberapa tahap yaitu *cleaning*, *case folding*, *tokenizing*, *normalisasi*, *stopword removal* dan *stemming*. Tahap ini berhasil memperoleh 1000 data yang menghasilkan 910 data yang berhasil lolos dari tahap *preprocessing*.

Cleaning

Dalam tahapan ini, dilakukan proses menghapus elemen-elemen khusus, angka, URL, dan tanda baca yang tidak diperlukan dalam analisis sentimen. Berikut adalah elemen-elemen yang dihapus dari data.

```
# 1. Cleaning
# Hapus URL
text = re.sub(r'http\S+|www\S+|https\S+', '', text)

# Hapus mention (@username)
text = re.sub(r'@\w+', '', text)

# Hapus seluruh hashtag (#hashtag)
text = re.sub(r'#\w+', '', text)

# Hapus RT (retweet)
text = re.sub(r'RT\S+', '', text)

# Hapus karakter non-ASCII
text = ''.join(c for c in text if ord(c) < 128)
```

Gambar 4. Script *cleaning*

Elemen-elemen ini, dihilangkan dengan tujuan untuk membuat kalimat menjadi bersih. Hasil *cleaning* diperlihatkan pada tabel 1.

Tabel 1.
Hasil *cleaning data*

Sebelum	Sesudah
@ImaLahmuddin @tikalumis @GasMedsos66731 Lah gue bukan dari kelompok yg menyuarakan #kaburajadulu. Salah orang loe.	Lah gue bukan dari kelompok yg menyuarakan #kaburajadulu. Salah orang loe..

Case folding

Dalam tahapan ini, dilakukan proses mengubah seluruh teks menjadi huruf kecil (*lowercase*) untuk menyamakan representasi kata. Misalnya, kata “Setuju” dan “setuju” dianggap sama setelah proses ini.

```
# 2. Case Folding
text = text.lower()
```

Gambar 5. Script case folding

Tahapan ini dilakukan dengan mengonversi huruf besar dari perubahan ukuran huruf. Hasil diperlihatkan pada tabel 2.

Tabel 2.
Hasil *case folding*

Sebelum	Sesudah
Lah gue bukan dari kelompok yg menyuarakan #kaburajadulu. Salah orang loe..	lah gue bukan dari kelompok yg menyuarakan #kaburajadulu. salah orang loe..

Normalization

Dalam tahapan ini, dilakukan proses mengubah kata-kata tidak baku atau slang menjadi bentuk yang standar, seperti “yg” menjadi “yang”, atau “sy” menjadi “saya”. Hasil diperlihatkan pada tabel 3.

```
# 7. Normalisasi Kata Slang
tokens = [combined_slang_dict.get(token, token) for token in tokens]
```

Gambar 6. Script normalization

```
# Normalisasi Bahasa Indonesia (Singkatan Umum)

norm = {
    " yg ": " yang ",
    " nggak ": " tidak ",
    " gak ": " tidak ",
    " bangetdari ": " banget dari ",
    " vibes ": " suasana ",
    " sm ": " sama ",
    " dr ": " dari ",
    " dgn ": " dengan ",
    " kl ": " kalau ",
```

Gambar 7. Kata-kata dalam *normalization*

Tabel 3.
Hasil *normalization*

Sebelum	Sesudah
lah gue bukan dari kelompok yg menyuarakan #kaburajadulu. salah orang loe..	lah gue bukan dari kelompok yang menyuarakan #kaburajadulu. salah orang loe..

Stopword Removal

Dalam tahapan ini, dilakukan proses menghilangkan kata-kata umum seperti “yang” yang tidak memiliki makna penting dalam konteks sentimen. Hasil diperlihatkan pada tabel 4..

```
# 8. Stopword Removal
tokens = [token for token in tokens if token not in all_stopwords]
```

Gambar 8. Script *stopword removal*Tabel 4.
Hasil *stopword removal*

Sebelum	Sesudah
lah gue bukan dari kelompok yg menyuarakan #kaburajadulu. salah orang loe..	lah gue bukan dari kelompok menyuarakan #kaburajadulu. salah orang lu..

Tokenizing

Dalam tahapan ini, dilakukan proses memisahkan kalimat menjadi potongan-potongan kata (token) yang merupakan unit analisis dalam pemrosesan teks. Hasil diperlihatkan pada tabel 5.

```
# 6. Tokenizing
tokens = word_tokenize(text)
```

Gambar 9. Script *tokenizing*Tabel 5.
Hasil *tokenizing*

Sebelum	Sesudah
lah gue bukan dari kelompok menyuarakan #kaburajadulu. salah orang lu..	[lah, gue, bukan, dari, kelompok, menyuarakan, #kaburajadulu, salah, orang, lu..]

Stemming

Dalam tahapan ini, dilakukan proses mengubah kata ke bentuk dasar atau akar katanya, misalnya “berlari”, “lari-lari” dan “menyuarakan” menjadi “suara”. Hasil *stemming* diperlihatkan pada tabel 6.

```
# 9. Stemming
stemmed_tokens = [stemmer.stem(token) for token in tokens]

# Gabungkan kembali menjadi teks
cleaned_text = ' '.join(stemmed_tokens)
```

Gambar 10. Script *stemming*Tabel 6.
Hasil *stemming*

Sebelum	Sesudah
[lah, gue, bukan, dari, kelompok, menyuarakan, #kaburajadulu, salah, orang, lu..]	lah gue bukan dari kelompok suara #kaburajadulu salah orang lu..

Labeling

Tahap selanjutnya pemberian label dengan menentukan kata-kata yang memiliki arti positif, netral dan negatif. Jumlah kata dari sebuah kalimat pada dataset akan dihitung sesuai indikator. Terdapat 3 jenis indikator label yang digunakan, indikator positif jika polaritas antara 0.0 hingga 1.0, indikator netral jika polaritas sama dengan 0.0, dan indikator negatif jika polaritas antara 0.0 hingga -1.0. Penetapan nilai polaritas dilakukan menggunakan *Python* yang dijalankan dengan *Google Colab*. Hasil *Labeling* dapat diperlihatkan pada tabel 7.

Tabel 7.
Hasil *labeling*

Dataset	Nilai Sentimen	Hasil Sentimen
nih thread buat yang mau #kaburajadulu ke sg mungkin beberapa bisa membantu karna sekarang udah banyak perubahan juga... yang penting niat kalian bulat dan yang terpenting mental kalian harus mantap dulu... menyesuaikan dengan <i>fast paced</i> nya orang sg agak riweh... tapi kalo kalian mau	1	positif
saya dah hilang harapan #kaburajadulu	-1	negatif
langkah awal #kaburajadulu <i>not good enough</i> tapi okelah <i>for self taught</i> + gaada belajar	0	netral

Pembobotan Kata

Pada tahap ini dilakukan pembobotan kata dengan *term frequency-inverse document frequency* (TF-IDF) yang dapat mengukur dan mengekstrak tingkat relativitas dari kata-kata yang muncul dalam sebuah komentar. Semakin banyak kata yang berulang pada data, semakin tinggi nilai bobotnya. Dalam penelitian ini peneliti menghitung pembobotan TF-IDF melalui cara perhitungan serta pengujian menggunakan *script python* pada *google colab*. Berikut adalah pembobotan kata TF-IDF dilihat pada gambar 11.

```
df_clean = df_sentiment[df_sentiment['cleaned_text'] != ''].copy()
df_clean = df_clean.drop_duplicates(subset=['cleaned_text', 'sentiment'], keep='first')
print(f"Jumlah data setelah menghapus teks kosong: {len(df_clean)}")
print(f"Jumlah data setelah menghapus duplikat: {len(df_clean)}")

X = df_clean['cleaned_text']
y = df_clean['sentiment']
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3, random_state=42, stratify=y)

# TF-IDF tanpa n-gram
tfidf_vectorizer = TfidfVectorizer(max_features=5000)
X_train_tfidf = tfidf_vectorizer.fit_transform(X_train)
X_test_tfidf = tfidf_vectorizer.transform(X_test)
print(f"Dimensi TF-IDF untuk data training: {X_train_tfidf.shape}")

# Konversi untuk GaussianNB (dense array)
X_train_dense = X_train_tfidf.toarray()
X_test_dense = X_test_tfidf.toarray()
```

Jumlah data setelah menghapus teks kosong: 840
 Jumlah data setelah menghapus duplikat: 840
 Dimensi TF-IDF untuk data training: (588, 2093)

Gambar 11. *Script* pembobotan kata

Klasifikasi Naïve Bayes Classifier

Selanjutnya ialah perhitungan klasifikasi 910 data menggunakan algoritma *Naïve Bayes Classifier*. Klasifikasi dilakukan dengan menggunakan metode split data yaitu data latih diterapkan dengan 80% dan 20% sebagai data uji. Penggunaan perbandingan 80:20 membantu memaksimalkan penggunaan data untuk pelatihan dan menyediakan subset data yang cukup besar untuk evaluasi kinerja model. Jenis klasifikasi yang digunakan yaitu MultinomialNB untuk mendapatkan hasil terbaik. MultinomialNB digunakan untuk data diskrit, seperti jumlah kemunculan kata dalam dokumen. Hasil *script* menggunakan *google colab* dapat diperlihatkan pada gambar 12.

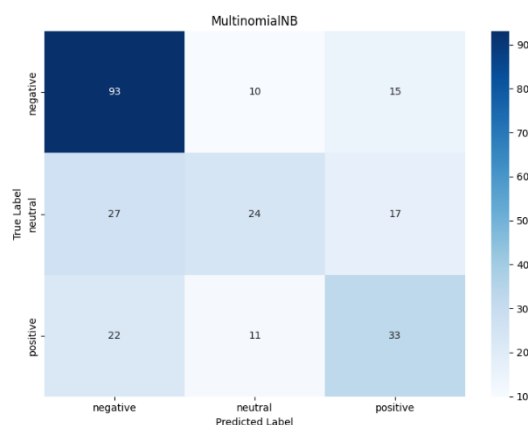
```
[ ] # Fungsi untuk menampilkan confusion matrix
def plot_confusion_matrix(y_true, y_pred, title, labels):
    conf_matrix = confusion_matrix(y_true, y_pred)
    plt.figure(figsize=(8, 6))
    sns.heatmap(conf_matrix, annot=True, fmt='d', cmap='Blues',
                xticklabels=labels, yticklabels=labels)
    plt.title(title)
    plt.xlabel('Predicted Label')
    plt.ylabel('True Label')
    plt.tight_layout()
    plt.savefig(f'confusion_matrix_{title.lower().replace(" ", "_")}.png')
    plt.show()

# Inisialisasi list untuk menyimpan hasil evaluasi
results = []
```

Gambar 12. Script klasifikasi Naïve Bayes Classifier

Evaluasi

Tahap selanjutnya setelah sistem klasifikasi yaitu pengujian terhadap sistem tersebut. Pengujian ini akan melihat sistem apakah sistem yang telah dibuat memiliki akurasi yang cukup tinggi, sehingga layak digunakan. *Confusion matrix* biasanya yang digunakan dalam pengujian akurasi ini, bukan hanya akurasi tapi *confusion matrix* bisa menguji presisi, recall dan f1-score. Dalam *confusion matrix* terdapat 4 rumus yaitu : *true positive*, *true negative*, *false positive* dan *false negative*. Metode *confusion matrix* diperlihatkan pada gambar 13.



Gambar 13. Confusion matrix

Tahap evaluasi, mendapatkan nilai hasil klasifikasi dari pengujian setiap perbandingan pada data latih dan data uji menggunakan *google colab*, yang menghasilkan nilai *confusion matrix* dari setiap data label. Hasil klasifikasi *Naïve Bayes Classifier* menghasilkan *accuracy* sebesar 56%, berdasarkan evaluasi dengan metode *confusion matrix*. Hasil klasifikasi *Naïve Bayes Classifier* setiap kelas diperlihatkan pada tabel 8.

Tabel 8.
Hasil klasifikasi *Naïve Bayes Classifier*

Label	Precision	Recall	F1-Score	Accuracy
Positif	0.51	0.50	0.50	0.56
Netral	0.53	0.35	0.42	
Negatif	0.65	0.79	0.72	

Label positif memiliki tingkat presision yang baik (50%) yang menunjukkan sebagian tweet diklasifikasikan konsisten sebagai positif, tingkat *recall* sebesar (50%) menunjukkan bahwa sentimen terdeteksi berhasil diklasifikasikan sebagai positif, meskipun ada beberapa sentimen positif yang terlewatkan dan *F1-Score* sebesar (50%) menunjukkan keseimbangan yang baik antara presicion dan *recall*, label netral menunjukkan performa terendah dengan nilai *recall* hanya sebesar (35%) dan *F1-score* sebesar (42%) meskipun nilai *precision* (53%). Hal ini menunjukkan bahwa model masih mengalami kesulitan dalam membedakan data

Visualisasi

[illegible]

Volume 10 Nomor 2 September 2025, **DAYA – MAS** | 102



Hasil ini menunjukkan bahwa dari ketiga variabel yang diuji, hanya variabel "Kata-Kata dalam Tweet" yang secara signifikan mempengaruhi model regresi multinomial.

Hasil penelitian menunjukkan bahwa melalui tweet dipengaruhi oleh komunikasi interpersonal dan komunikasi massa. Komunikasi interpersonal memiliki daya pengaruh tinggi karena memberikan ruang keterlibatan emosional dan interaksi dua arah antara pelaku pengguna Twitter. Sementara media sosial, meskipun merupakan bentuk komunikasi massa digital, mengandung elemen-elemen interpersonal melalui fitur interaksi seperti komentar, balasan, dan penggunaan bahasa yang personal. Oleh karena itu, penggunaan kata-kata dalam tweet dapat dilihat sebagai upaya untuk menjembatani komunikasi massa dengan pendekatan interpersonal, menjadikan media sosial sebagai ruang hibrida yang menampung keduanya.

Pada temuan ini relevan dengan melalui assembly theory, media memiliki peran penting dalam membentuk realitas sosial dan persepsi publik melalui fenomena isu tertentu dan cara penyajiannya. Kata-kata dalam tweet berperan sebagai penyusun agenda simbolik yang dapat memengaruhi pembentukan opini publik. Media sosial memungkinkan individu untuk menjadi pengguna penyusun narasi dan memilih kata-kata yang mampu membingkai isu secara efektif. Hal ini menjelaskan mengapa pemilihan kata dalam pesan digital memiliki kekuatan lebih besar dibandingkan elemen simbolik lainnya seperti emoji. Dengan kata lain, dalam era digital, kekuatan komunikasi interpersonal dan massa tidak hanya terletak pada saluran yang digunakan, tetapi juga pada bagaimana pesan dikonstruksi secara linguistik untuk menciptakan resonansi dan sosial di benak masyarakat sehingga menghasilkan sentimen.

Tabel 10.
Hasil Model Fitting Information

Model	Model Fitting Criteria	Likelihood Ratio Test		
		Chi-Square	Df.	Sig.
Intercept Only	309.789			
Final	188.583	121.206	96	.042

Berdasarkan hasil analisa model fitting information pada uji regresi multinomial, dapat diketahui bahwa nilai $-2 \text{ Log Likelihood}$ untuk model yang hanya menggunakan intersep (Intercept Only) adalah sebesar 309.789. Sementara itu, model akhir (Final) yang telah memasukkan seluruh variabel independen menunjukkan nilai $-2 \text{ Log Likelihood}$ sebesar 188.583. Selisih antara kedua nilai tersebut menghasilkan nilai Chi-Square sebesar 121.206 dengan derajat kebebasan (df) sebanyak 96 dan nilai signifikansi (Sig.) sebesar 0.042. Nilai signifikansi ini lebih kecil dari ambang batas 0.05, sehingga dapat disimpulkan bahwa model akhir secara statistik memberikan peningkatan yang signifikan dibandingkan dengan model dasar yang hanya menggunakan intersep. Dengan demikian, keberadaan variabel-variabel independen dalam model memberikan kontribusi yang berarti terhadap pemodelan variabel dependen, dan model dapat dianggap layak untuk digunakan dalam menjelaskan hubungan yang diuji.

Hasil penelitian menunjukkan bahwa struktur model yang dibangun bukan sekadar bersifat acak atau tidak bermakna, melainkan memiliki kekuatan prediktif yang nyata terhadap bentuk komunikasi yang dianalisis dalam penggunaan tagar #kaburajadulu di Twitter. Variable-variabel yang dimasukkan ke dalam model seperti kata-kata dalam tweet, penggunaan emoji, dan frekuensi kemunculan kata secara bersama-sama memberikan pengaruh terhadap cara pengguna membentuk dan menyampaikan pesan politik di ruang digital. Hal ini sekaligus memperkuat validitas model dalam konteks penelitian ini dan menandakan bahwa pendekatan kuantitatif yang digunakan dapat diandalkan dalam mengeksplorasi fenomena sosial dengan menganalisa komunikasi di Twitter.

Hasil ini juga mendukung bahwa pentingnya konten, struktur pesan, dan konteks linguistik dalam memengaruhi respons masyarakat. Model akhir yang signifikan

mengindikasikan bahwa interaksi antara faktor-faktor linguistik dalam media sosial berperan aktif dalam menentukan bagaimana sebuah pesan dipersepsikan dan bagaimana komunikasi itu terbentuk. Komunikasi interpersonal memperkuat gagasan bahwa pesan-pesan yang tersebar melalui media sosial tidak hanya tergantung pada platform atau saluran komunikasi semata, tetapi sangat dipengaruhi oleh umpan balik, substansi dan penyusunan bahasa. Hasil ini dapat dikatakan layak digunakan dan kelayakan model ini juga menunjukkan bahwa komunikasi di media sosial dapat dipetakan secara terstruktur dan dianalisis menggunakan pendekatan statistik regresi logistik multinomial.

KESIMPULAN

Penelitian ini memberikan temuan bahwa algoritma *Naïve Bayes Classifier* yang digunakan untuk mengklasifikasikan sentimen pada tweet dengan tagar #kaburajadulu menghasilkan tingkat akurasi sebesar 56%. Meskipun akurasi tersebut belum tergolong tinggi, model menunjukkan performa yang cukup baik dalam mengenali sentimen negatif, dengan nilai *precision* sebesar 65%, *recall* sebesar 79%, dan *F1-score* sebesar 72%. Ini mengindikasikan bahwa model relatif konsisten dan akurat dalam mendeteksi tweet yang bermuatan negatif. Di sisi lain, sentimen positif diklasifikasikan dengan performa yang moderat, masing-masing nilai *precision*, *recall*, dan *F1-score* sebesar 50%. Sedangkan sentimen netral menjadi kategori dengan performa terendah, terutama dari sisi *recall* yang hanya mencapai 35%, yang menandakan bahwa model sering gagal mengenali tweet yang sebenarnya netral, dan mengklasifikasikannya sebagai sentimen lain. Hal ini menunjukkan bahwa model masih mengalami kesulitan dalam membedakan nuansa netral dari positif dan negatif, yang turut berkontribusi pada rendahnya akurasi keseluruhan.

Hasil analisis regresi logistik multinomial terhadap bentuk komunikasi dalam tagar #kaburajadulu di Twitter menunjukkan bahwa terdapat variabel-variabel tertentu yang berpengaruh signifikan terhadap kecenderungan pengguna dalam menyampaikan pesan dengan gaya komunikasi yang berbeda, seperti ekspresif, ajakan, atau informatif. Model ini mampu mengidentifikasi keterkaitan antara karakteristik pengguna dan bentuk komunikasi yang dipilih, yang mengindikasikan bahwa perilaku komunikasi di media sosial dapat dipetakan secara sistematis menggunakan pendekatan statistik.

Kelebihan penelitian ini terletak pada integrasi antara analisis sentimen dan analisis bentuk komunikasi, yang memberikan gambaran menyeluruh mengenai cara pengguna berekspresi dalam merespons isu sosial di media sosial. Namun, kekurangan yang ditemukan adalah keterbatasan dalam akurasi klasifikasi, terutama untuk sentimen netral, serta keterbatasan dalam representasi fitur teks yang digunakan, yang masih bersifat dasar.

Berdasarkan temuan penelitian, analisis bentuk komunikasi masih dapat dikembangkan variabel-variabel kontekstual tambahan seperti waktu, interaksi sosial, atau jaringan percakapan, agar pemetaan pola komunikasi menjadi lebih akurat dan kontekstual.

UCAPAN TERIMA KASIH

Penulis mengucapkan terimakasih yang sebesar-besarnya kepada semua pihak yang telah memberikan dukungan dalam pelaksanaan penelitian ini. Ucapan terimakasih khusus disampaikan kepada dosen pembimbing atas bimbingan dan arahnya. Tanpa bantuan dan dukungan tersebut, penelitian ini tidak akan dapat terselesaikan dengan baik.

DAFTAR PUSTAKA

- Abdullah RN, Ramadhan AM. (2024). Analisis Komunikasi Interpersonal Pemain Korean Roleplay di Lingkungan Keluarga. *Al-Itishol: Jurnal Komunikasi dan Penyiaran Islam*. 5 (2), 171-186
- Agustin AV, Sa'adah FN, & Umaidah Y. (2024). Analisis Sentimen Menggunakan Metode *Naive Bayes* Terhadap *Childfree*. *Dinamik*.

- Alam S, Sulisty MI. (2023). Analisis Sentimen Berdasarkan Ulasan Pengguna Aplikasi MyPertamina Pada Google Playstore Menggunakan Metode Naïve Bayes. *Storage: Jurnal Ilmiah Teknik Dan Ilmu Komputer*. 2 (3), 100-108
- Amrullah AZ, Anas AS, Hidayat MAJ. (2020). Analisis Sentimen Movie Review Menggunakan Naive Bayes Classifier Dengan Seleksi Fitur Chi Square. *Jurnal Bumigora Information Technology (BITE)*. 2 (1), 40-44
- Andhika A, Wihardi D. (2022). Pembungkahan Berita Pencopotan Kapolda Metro Jaya Dan Kapolda Jawa Barat Di Portal Berita Online Kompas. Com Dan Detik. Com Periode 16-18 November 2020. *PANTAREI*.
- Anggraini C, Ritonga DH, Kristina L, Syam M, Kustiawan W. (2022). Komunikasi interpersonal. *Jurnal Multidisiplin Dehasen (MUDE)*. 1(3), 337-342.
- Artanto FA (2024). Analisis Sentimen Opini Publik terhadap Fenomena Bunuh Diri Mahasiswa di Twitter Menggunakan Algoritma Naive Bayes. *SATESI: Jurnal Sains Teknologi dan Sistem Informasi*. 4 (1), 70-77
- Aryanti PG, Santoso I. (2023). Analisis Sentimen Pada Twitter Terhadap Mobil Listrik Menggunakan Algoritma Naive Bayes. *IKRA-ITH Informatika: Jurnal Komputer Dan Informatika*. 7 (2), 133-137
- Asosiasi Penyelenggara Jasa Internet Indonesia (APJII). (2024). APJII Jumlah Pengguna Internet Indonesia Tembus 221 Juta Orang.
- Djamaludin, M. A., Triayudi, A., & Mardiani, E. (2022). Analisis sentimen tweet kri nanggala 402 di twitter menggunakan metode naïve bayes classifier. *Jurnal JTIK (Jurnal Teknologi Informasi dan Komunikasi)*. 6 (2), 161-166
- Fachlevi S, Unde AA, Hasrullah H. (2024). Analisis Sentimen Publik pada Tagar# KawalPutusanMK di Media Sosial X. *Jurnal Syntax Admiration*. 5(10), 4232-4241
- Fauziah R, Zulkamain I. (2023). Pola Komunikasi Relawan Digital pada Akun Twitter dan Instagram Sekolah Relawan terhadap Manajemen Bencana di Indonesia. *Journal Of Science And Social Research*. 6 (1) 60-65
- Halim A, Safuwani A. (2023). Analisis Sentimen Opini Warganet Twitter Terhadap Tes Screening Genose Pendeteksi Virus Covid-19 Menggunakan Metode Naïve Bayes Berbasis Particle Swarm Optimization. *Jurnal Informatika Teknologi dan Sains (Jinteks)*. 5 (1), 170-178
- Hanum AJ. (2025). Efektivitas Komunikasi Interpersonal Dr. Zaldi Sp. M Dalam Mengedukasi Pasien Katarak Di Rsu Citra Medika (Doctoral dissertation, Fakultas Ilmu Sosial dan Ilmu Politik, Universitas Islam Sumatera Utara).
- Hermawanti F, Prisanto GF, Yulianto K, Ruliana P. (2020). Pengaruh Terpaan Media# GundikLintasBUMN Pada Twitter terhadap persepsi profesi pramugari. *Prosiding Hubungan Masyarakat*.
- Kevin K, Enjeli M, Wijaya A. (2024). Analisis Sentimen Penggunaan Aplikasi Kinemaster Menggunakan Metode Naive Bayes. *Jurnal Ilmiah Computer Science*. 2 (2), 89-98
- Lestari RD, Nurisulawati I. (2024). Analisis Sentimen Masyarakat terhadap Kampanye Sosial Pengurangan Food Waste di Indonesia. *Uranus: Jurnal Ilmiah Teknik Elektro, Sains dan Informatika*. 2(2): 15-26.
- Mustaqillah R., Widyaningtyas O, Wantoro T. (2023). Efektivitas Penggunaan Twitter Sebagai Sarana Peningkatan Berpikir Kritis Mahasiswa Ilmu Komunikasi. *MUKASI: Jurnal Ilmu Komunikasi*. 2(1), 18-28.
- Nurtikasari Y, Alam S, Hermanto TI. (2022). Analisis Sentimen Opini Masyarakat Terhadap Film Pada Platform Twitter Menggunakan Algoritma Naive Bayes. *INSOLOGI: Jurnal Sains dan Teknologi*. 1 (4), 411-423
- Rahmawati Y, Hariyati F, Abdullah AZ, Nurmianani M. (2024). Gaya komunikasi dakwah era digital: Kajian literatur. *Concept: Journal of Social Humanities and Education*. 3 (1), 266-279
- Rakhman FR, Ramadhani RW, Kuncoroyakti YA. (2021). Analisis Sentimen Dan Opini Digital Kampanye 3m Di Masa Covid-19 Melalui Media Sosial Twitter. *KOMUNIKOLOGI: Jurnal Ilmiah Ilmu Komunikasi*. 18 (01) 8-21
- Putra SA, Wijaya A. (2023). Analisis Sentimen Artificial Intelligence (Ai) Pada Media Sosial Twitter Menggunakan Metode Lexicon Based. *JuSiTik: Jurnal Sistem dan Teknologi Informasi Komunikasi*. 7 (1), 21-28
- Putri KS, Setiawan IR, Pambudi A. (2023). Analisis Sentimen Terhadap Brand Skincare Lokal Menggunakan Naïve Bayes Classifier. *Technologia: Jurnal Ilmiah*. 14 (3), 227-232
- Saleh R. (2024). Analisis Komunikasi Pendidikan Pada Pengembangan Konten Pembelajaran Smart School Dinas Pendidikan Provinsi Sulawesi Selatan. *Jurnal Komunikasi*. 17 (1), 37-46
- Tarigan FA, Robin R, Angeline S. (2024). Pengaruh Digital Marketing Dan Brand Awareness Terhadap Keputusan Pembelian Pada Produk Aqua Di Desa Cinta Rakyat. *Jurnal Ilmiah Manajemen, Ekonomi, & Akuntansi (MEA)*. 8 (2), 326-336