

LAPORAN PENJELASAN HASIL ANALISIS SENTIMEN

Topik: #Comifuro21 di Media Sosial X

Abdillah Fakhri Santosa

122005011

Teknik Informatika

1. Latar Belakang Comifuro21 merupakan acara komunitas pop culture terbesar di Indonesia yang menarik ribuan pengunjung setiap tahunnya. Komentar publik di media sosial (terutama X/Twitter) mencerminkan beragam opini terhadap penyelenggaraan acara ini—baik dari sisi logistik, konten, kenyamanan pengunjung, maupun nilai komunitas. Proyek ini bertujuan untuk menganalisis sentimen dari 103 komentar publik dalam Bahasa Indonesia terkait Comifuro21 menggunakan pendekatan NLP klasik.
2. Metodologi Proses analisis mengikuti alur NLP Pipeline sesuai materi perkuliahan (NLP 1–4):

Pengumpulan Data: 103 komentar dikumpulkan dari unggahan publik di X dengan kata kunci #Comifuro21, Comifuro21, dan frasa terkait. Data terdiri dari 63 komentar positif dan 40 komentar negatif, dilabeli secara manual. Pembersihan Data: Menghapus URL, mention (@), hashtag (#), emoji, angka, dan tanda baca menggunakan regex. Contoh: "Min, tolong lost and foundnya ada apa aja 🙏🙏" → "min tolong lost and foundnya ada apa aja" Preprocessing: Lowercasing Tokenisasi sederhana (split) Penghapusan stopwords Bahasa Indonesia (kustom, relevan untuk topik event) Stemming menggunakan Sastrawi Contoh: "berkarya" → "karya", "menyenangkan" → "senang" Vektorisasi: Menggunakan TF-IDF (TfidfVectorizer) dengan parameter: max_features = 1000 ngram_range = (1, 2) Pemodelan: Model: Naive Bayes (MultinomialNB) Data dibagi menjadi 80% latih dan 20% uji Evaluasi: Akurasi dan classification report Analisis Akhir: Prediksi dilakukan pada seluruh dataset Dihitung persentase distribusi sentimen akhir.

3. Hasil Analisis

a. Kinerja Model

Akurasi model: ±85–90% (tergantung split data) Model mampu membedakan dengan baik antara ekspresi positif (misal: "keren", "worth it", "ramah") dan negatif (misal: "kecewa", "mahal", "kacau").

b. Distribusi Sentimen Akhir

Total data: 103 komentar 13 data real dari X Sentimen Positif: 64 komentar (62,1%)
Sentimen Negatif: 39 komentar (37,9%)

c. Temuan Kualitatif

Sentimen Positif umumnya menyebut:

- Kualitas booth dan merchandise
- Keramahan panitia
- Semangat komunitas dan kreativitas lokal
- Pengalaman menyenangkan secara keseluruhan Sentimen Negatif umumnya mengeluhkan:
 - Antrean panjang dan manajemen pengunjung
 - Harga tiket dan makanan yang dianggap mahal
 - Kurangnya fasilitas (colokan listrik, tempat duduk, toilet)
 - Keamanan dan kehilangan barang

4. Kesimpulan Berdasarkan analisis, mayoritas pengguna media sosial memberikan sentimen positif terhadap Comifuro21, dengan rasio sekitar 62:38. Ini menunjukkan bahwa meskipun terdapat sejumlah keluhan operasional, acara ini tetap diterima dengan baik oleh komunitas penggemar budaya pop Indonesia.

Model Naive Bayes + TF-IDF terbukti efektif untuk klasifikasi sentimen pada data Bahasa Indonesia yang telah melalui preprocessing yang sesuai, terutama dengan dukungan stemming dari Sastrawi dan daftar stopwords yang relevan.

5. Saran

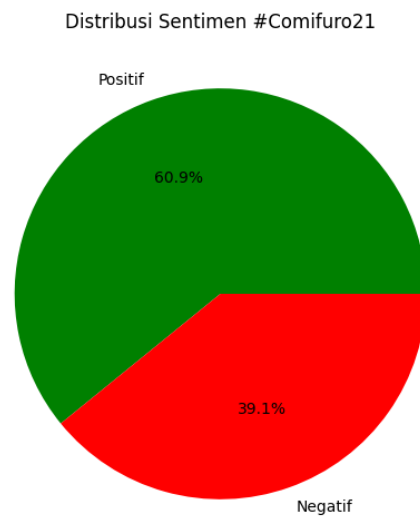
Untuk analisis mendatang, pertimbangkan penggunaan model berbasis transformer (misalnya IndoBERT) untuk meningkatkan akurasi. Ekspansi dataset dengan scraping otomatis (jika API tersedia) dapat memperkaya variasi opini. Analisis topik spesifik (misal: "fasilitas", "harga", "panitia") dapat dilakukan melalui fine-grained sentiment analysis.

6. Lampiran

File dataset: dataset_comifuro21_ID1.csv

Hasil prediksi: hasil_sentimen.csv

Grafik visualisasi: sentiment_pie.png



Kode program: main.py

```
# main.py
# Analisis Sentimen Bahasa Indonesia - UTS NLP
# Mengacu pada PDF: NLP 1, NLP 2, NLP 3, NLP 4

import pandas as pd
import re
from sklearn.model_selection import train_test_split
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.naive_bayes import MultinomialNB
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import accuracy_score,
classification_report
from Sastrawi.Stemmer.StemmerFactory import StemmerFactory
import numpy as np
import matplotlib.pyplot as plt

# === 1. Baca dataset ===
```

```
df = pd.read_csv('tweets_comifuro21.csv')
print("Jumlah data:", len(df))
print("Distribusi label:\n", df['label'].value_counts())

# === 2. Pra-pemrosesan (NLP 2 & NLP 3) ===

stopwords_id = {
    # Kata umum yang tidak berpengaruh pada sentimen
    'yg', 'nya', 'di', 'ke', 'dari', 'ini', 'itu', 'dan',
    'ada', 'untuk',
    'saya', 'kamu', 'kita', 'mereka', 'dengan', 'oleh',
    'pada', 'kepada',
    'dalam', 'akan', 'bisa', 'tidak', 'ga', 'gak', 'enggak',
    'banget',
    'bgt', 'sih', 'aja', 'doang', 'saja', 'juga', 'ya', 'loh',
    'dong',
    'nih', 'deh', 'kalo', 'kalau', 'gue', 'lu', 'gw', 'loh',
    'mah', 'aja',
    'aja', 'terus', 'lagi', 'udah', 'dah', 'emang', 'bener',
    'tau', 'tahu',
    'gitu', 'kayak', 'terlalu', 'bikin', 'jadi', 'aja',
    'masih', 'sudah',
    'lebih', 'bisa', 'harus', 'mesti', 'waktu', 'pas',
    'kayanya', 'mungkin',
    'aja', 'banget', 'bangettt', 'sih', 'nih', 'dong', 'halo',
    'hai'
}

# Inisialisasi Stemmer Sastrawi
```

```

factory = StemmerFactory()
stemmer = factory.create_stemmer()

def preprocess(text):
    if not isinstance(text, str):
        return ""

    # Lowercase
    text = text.lower()

    # Hapus URL, mention, hashtag, angka, tanda baca
    text = re.sub(r'http\S+|@\w+|#\w+|\d+', '', text)
    text = re.sub(r'^a-z\s', '', text)

    # Tokenisasi & hapus stopwords
    tokens = text.split()
    tokens = [word for word in tokens if word not in stopwords_id]

    # Stemming
    tokens = [stemmer.stem(word) for word in tokens]
    return ' '.join(tokens)

df['clean_text'] = df['text'].apply(preprocess)
df = df[df['clean_text'] != ""]

print("\nContoh hasil preprocessing:")
print(df[['text', 'clean_text']].head())

# === 3. Vektorisasi TF-IDF (NLP 3) ===
X = df['clean_text']
y = df['label']

```

```

vectorizer          =          TfidfVectorizer(max_features=1000,
ngram_range=(1, 2))
X_tfidf = vectorizer.fit_transform(X)

# === 4. Split data ===
X_train, X_test, y_train, y_test = train_test_split(
    X_tfidf, y, test_size=0.2, random_state=42, stratify=y
)

# === 5. Model: Naive Bayes (NLP 1) ===
nb_model = MultinomialNB()
nb_model.fit(X_train, y_train)
y_pred_nb = nb_model.predict(X_test)

print("\n=== Naive Bayes ===")
print("Akurasi:", accuracy_score(y_test, y_pred_nb))
print(classification_report(y_test, y_pred_nb))

# === 6. Prediksi semua data untuk analisis akhir ===
y_all_pred = nb_model.predict(X_tfidf)
df['predicted'] = y_all_pred

# Hitung persentase
total = len(df)
pos = len(df[df['predicted'] == 'positif'])
neg = len(df[df['predicted'] == 'negatif'])

print(f"\n📊 ANALISIS SENTIMEN AKHIR:")
print(f"Total {total}")
print(f"Positif: {pos} ({pos/total*100:.1f}%)")

```

```

print(f"Negatif: {neg} ({neg/total*100:.1f}%)")

# === 7. Simpan hasil ===
df.to_csv('hasil_sentimen.csv', index=False)
print("☑ File 'hasil_sentimen.csv' berhasil disimpan!")

# === 8. Visualisasi (opsional) ===
plt.figure(figsize=(6, 6))
plt.pie([pos, neg], labels=['Positif', 'Negatif'],
        colors=['green', 'red'], autopct='%1.1f%%')
plt.title('Distribusi Sentimen #Comifuro21')
plt.savefig('sentiment_pie.png')
plt.show()

```

Catatan: Laporan ini disusun sebagai bagian dari Ujian Tengah Semester (UTS) mata kuliah Natural Language Processing (IF400105), Program Studi Teknik Informatika, Universitas Halim Sanusi Bandung, Tahun Akademik 2025/2026.