



**PROGRAM STUDI TEKNIK INFORMATIKA**

**FAKULTAS TEKNIK**

**UNIVERSITAS MURIA KUDUS**

**2021**

## HALAMAN PERSETUJUAN

### ANALISIS SENTIMEN PENGGUNA YOUTUBE TERHADAP TAYANGAN #MATANAJWA MENANTI TERAWAN DENGAN METODE NAIVE BAYES CLASSIFIER



## HALAMAN PENGESAHAN

### ANALISIS SENTIMEN PENGGUNA YOUTUBE TERHADAP TAYANGAN #MATANAJWAMENANTI TERAWAN DENGAN METODE NAIVE BAYES CLASSIFIER



## **PERNYATAAN KEASLIAN**

Saya yang bertanda tangan dibawah ini:

Nama : Mathoril Hudha  
NIM : 201651062  
Tempat & Tanggal Lahir : Kudus, 1 November 1998  
Judul Skripsi : Analisis Sentimen Pengguna *YouTube* terhadap Tayangan #MataNajwaMenantiTerawan dengan Metode *Naïve Bayes Classifier*

Menyatakan dengan sebenarnya bahwa penulisan Skripsi/Tugas Akhir\* ini berdasarkan hasil penelitian, pemikiran dan pemaparan asli dari saya sendiri, baik untuk naskah laporan maupun kegiatan lain yang tercantum sebagai bagian dari Skripsi ini. Seluruh ide, pendapat, atau materi dari sumber lain telah dikutip dalam Skripsi dengan cara penulisan referensi yang sesuai.

Demikian pernyataan ini saya buat dengan sesungguhnya dan apabila di kemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka saya bersedia menerima sanksi akademik berupa pencabutan gelar dan sanksi lain sesuai dengan peraturan yang berlaku di Universitas Muria Kudus.

Demikian pernyataan ini saya buat dalam keadaan sadar tanpa paksaan dari pihak manapun.

Kudus, 01 Januari 2022

Yang memberi pernyataan,



Mathoril Hudha

NIM. 201651062

# **ANALISIS SENTIMEN PENGGUNA YOUTUBE TERHADAP TAYANGAN #MATANAJWAMENANTITERAWAN DENGAN METODE NAÏVE BAYES CLASSIFIER**

Nama Mahasiswa : Mathoril Hudha

NIM : 201651062

Pembimbing :

1. Endang Supriyati S.Kom., M.Kom

2. Tri Listyorini S.Kom., M.Kom

## **RINGKASAN**

Tayangan #MataNajwaMenantiTerawan menuai pro dan kontra setelah Najwa Shihab dalam acara Mata Najwa mewawancara kursi kosong yang direpresentasikan sebagai Menteri Kesehatan Terawan Agus Putranto. Wawancara dengan kursi kosong biasa dilakukan di luar negeri yang memiliki histori pers yang lama. Seperti yang dilakukan Piers Morgan di CNN dan Kay Burley di acara Kay Burley Show di Sky News. Media sosial mengalami peningkatan penggunaan sehingga analisis sentimen perlu dilakukan untuk mengetahui opini publik terhadap isu yang sedang terjadi. Analisis sentimen dalam penelitian ini dilakukan untuk mengetahui bagaimana penilaian publik terhadap tayangan #MataNajwaMenantiTerawan di *Youtube*. Data dalam penelitian ini diambil dengan cara *scrapping* dari komentar video youtube #MataNajwaMenantiTerawan. Dataset mentah hasil *scrapping* dilakukan *preprocessing* untuk memudahkan proses klasifikasi. Algoritma *Naïve Bayes* digunakan untuk mengklasifikasikan komentar ke kategori sentimen positif, sentimen negatif dan netral. Algoritma *Naïve Bayes* menghasilkan akurasi sebesar 90.36%. Sentimen netral lebih mendominasi dengan jumlah 1232 data netral, 90 data negatif dan 78 data positif.

Kata kunci: Analisis Sentimen, *Naive Bayes*, *Text Preprocessing*,  
#MataNajwaMenantiTerawan



**SENTIMENT ANALYSIS OF USER COMMENTS ON  
YOUTUBE VIDEO #MATANAJWAMENANTITERAWAN  
WITH NAIVE BAYES CLASSIFIER METHOD**

*Student Name* : Mathoril Hudha

*Student Identity Number* : 201651062

*Supervisor* :

1. Endang Supriyati S.Kom., M.Kom
2. Tri Listyorini S.Kom., M.Kom

**ABSTRACT**

*The show #MataNajwaMenantiTerawan reaped pros and cons after Najwa Shihab in the Mata Najwa program interviewed an empty chair represented as Minister of Health Terawan Agus Putranto. Interviews with empty chair are usually conducted abroad, which have a long press history. Like Piers Morgan on CNN and Kay Burley on the Kay Burley Show on Sky News. Social media has increased its use so that sentiment analysis needs to be done to find out public opinion on the issues that are happening. Sentiment analysis in this study was conducted to find out how the public evaluates the #MataNajwaMenantiTerawan show on Youtube. The data in this study were taken by scrapping from comments on the youtube video #MataNajwaMenantiTerawan. The raw dataset from scrapping is preprocessed to facilitate the classification process. Naive Bayes algorithm is used to classify comments into categories of positive sentiment, negative sentiment and neutral. The Naive Bayes algorithm produces an accuracy of 90.36%. Neutral sentiment dominates with 1232 neutral data, 90 negative data and 78 positive data.*

*Keywords:* Sentiment Analysis, Naive Bayes, Text Preprocessing, #MataNajwaMenantiTerawan.

## KATA PENGANTAR

Puji dan syukur alhamdullillah penulis panjatkan kehadiran Allah SWT yang Maha melimpahkan kasih dan sayangnya sehingga penulisan skripsi untuk memperoleh gelar Sarjana Komputer dapat terselesaikan.

Penyusunan Skripsi dengan judul "Analisis Sentimen Pengguna YouTube terhadap Tayangan #MataNajwaMenantiTerawan" ini tak lepas dari bantuan dan dukungan beberapa pihak, untuk itu penulis menyampaikan ucapan terima kasih kepada:

1. Bapak Prof. Dr. Ir. Darsono, M.Si selaku rektor Universitas Muria Kudus
2. Ibu Endang Supriyati S. Kom., M.Kom dan Ibu Tri Listyorini S.Kom., M.Kom selaku dosen pembimbing.
3. Bapak Arief Susanto., M.Kom selaku dosen wali saya.
4. Seluruh dosen Teknik Informatika yang telah memberikan keilmuan kepada penulis selama masa studi.

Dalam penulisan skripsi ini tentunya terdapat kekurangan, oleh karena itu penulis menerima kritik dan saran agar penulis dapat lebih baik di masa mendatang. Semoga skripsi ini bermanfaat bagi pembacanya.

Kudus, 03 Januari 2022

Penulis

## DAFTAR ISI

HALAMAN PERSETUJUAN.....	ii
HALAMAN PENGESAHAN.....	iii
PERNYATAAN KEASLIAN.....	iv
RINGKASAN .....	v
<i>ABSTRACT</i> .....	vii
KATA PENGANTAR .....	viii
DAFTAR ISI.....	ix
DAFTAR GAMBAR .....	xii
DAFTAR TABEL.....	xiii
DAFTAR LAMPIRAN.....	xiv
DAFTAR ISTILAH DAN SINGKATAN .....	xv
BAB I PENDAHULUAN .....	1
1.1. Latar Belakang .....	1
1.2. Perumusan Masalah.....	2
1.3. Batasan Masalah.....	3
1.4. Tujuan.....	3
1.5. Manfaat.....	3
BAB II TINJAUAN PUSTAKA.....	4
2.1. Penelitian Terdahulu.....	4
2.2. Analisis Sentimen.....	5
2.3. Tayangan #MataNajwaMenantiTerawan .....	6
2.4. <i>YouTube</i> .....	7
2.5. Diagram Alir.....	7
2.6. <i>Scraping</i> .....	8
2.7. <i>Term Frequency</i> .....	9
2.8. <i>Naïve Bayes</i> .....	9
2.9. <i>Supervised Learning</i> .....	10
BAB III METODOLOGI.....	11
3.1. Objek Penelitian.....	11

3.2. Kebutuhan Sistem .....	11
3.3. Studi Literatur.....	11
3.3.1. <i>Time</i> .....	11
3.3.2. <i>Selenium</i> .....	12
3.3.3. Analisa Sistem .....	12
3.3.4. Python .....	12
3.3.5. NLTK .....	12
3.3.6. Scikit-Learn .....	12
3.4. Diagram Alur Proses .....	13
3.5. <i>Scrapping</i> .....	13
3.6. <i>Text Preprocessing</i> .....	14
3.6.1. <i>Cleansing</i> .....	15
3.6.2. <i>Case Folding</i> .....	15
3.6.3. <i>Tokenize</i> .....	15
3.6.4. <i>Stopword Removal</i> .....	15
3.6.5. <i>Stemming</i> .....	15
3.7. Algoritma <i>Naïve Bayes</i> .....	15
BAB IV HASIL DAN PEMBAHASAN .....	17
4.1. Implementasi .....	17
4.1.1. Pengumpulan Data.....	17
4.1.2. <i>Text Preprocesing</i> .....	18
4.1.3. Labelling .....	33
4.1.3. Vektorisasi .....	37
4.1.4. Klasifikasi <i>Naïve Bayes</i> .....	38
4.2. Hasil Pembahasan .....	39
4.2.1. Hasil <i>Scrapping</i> .....	39
4.2.2. Hasil Preprocessing.....	45
4.2.3. Hasil Labeling.....	46
4.2.4. Hasil Klasifikasi.....	49
4.2.5. <i>Model Predic</i> .....	51
BAB V PENUTUP.....	52
5.1. Kesimpulan.....	52

5.2. Saran .....	52
DAFTAR PUSTAKA .....	53
LAMPIRAN 1 .....	56
LAMPIRAN 2 .....	59
BIODATA PENULIS .....	62



## DAFTAR GAMBAR

Gambar 3.2. Alur Pengambilan Data .....	14
Gambar 3.1. Diagram Alur Proses .....	13
Gambar 3.3. <i>Text Preprocessing</i> .....	15
Gambar 3.4. Klasifikasi <i>Naïve Bayes</i> .....	16
Gambar 4.5. Program <i>Scraping</i> .....	17
Gambar 4.6. Tampilan Googlesheet dengan add-ons LogicSheet .....	18
Gambar 4.7. Program Vektorisasi.....	38
Gambar 4.8. Script Klasifikasi <i>Naïve Bayes</i> dengan <i>Library Scikit-Learn</i> .....	39
Gambar 4.9. Output Program <i>Scraping</i> .....	40
Gambar 4.10. Hasil <i>Preprocessing</i> .....	45
Gambar 4.11. Perhitungan <i>Term Frequency</i> .....	47
Gambar 4.12. Data yang diberi label .....	47
Gambar 4.13. Hasil Klasifikasi .....	50
Gambar 4.14. Prediksi Model .....	51

## DAFTAR TABEL

Tabel 2.1. Penghargaan Acara Mata Najwa dari Tahun 2010-2014 .....	6
Tabel 2.2. Simbol, Gambar dan Fungsinya.....	7
Tabel 4.3. <i>Cleansing</i> .....	18
Tabel 4.4. <i>Case Folding</i> .....	21
Tabel 4.5. <i>Tokenizing</i> .....	24
Tabel 4.6. <i>Stopword Removal</i> .....	27
Tabel 4.7. <i>Stemming</i> .....	30
Tabel 4.8. Perhitungan Term Positif .....	33
Tabel 4.9. Perhitungan Term Negatif .....	34
Tabel 4.10. Perhitungan Polaritas .....	36
Tabel 4.11. Hasil <i>Scrapping</i> .....	40
Tabel 4.12. Hasil Text Preprocessing .....	45
Tabel 4.13. Hasil Labelling .....	47
Tabel 4.14. Hasil Klasifikasi Data Training .....	50
Tabel 4.15. Hasil Klasifikasi Data Testing .....	50
Tabel 4.16. Hasil Klasifikasi Sentimen .....	50

## **DAFTAR LAMPIRAN**

Lampiran 1. Lembar Konsultasi .....	56
Lampiran 2. Lembar Revisi.....	59



## DAFTAR ISTILAH DAN SINGKATAN

NBC: *Naive Bayes Classifier*

IoT: *Internet of Things*

TF: *Term Frequency*

Covid-19: *Corona Virus Disease – 2019*

NLP: *Natural Language Processing*

NLTK: *Natural Language Toolkit*

