

**DETEKSI CLICKBAIT BERDASARKAN KOMENTAR PADA VIDEO
TRENDING DI YOUTUBE INDONESIA MENGGUNAKAN
NATURAL LANGUAGE PROCESSING**

SKRIPSI



disusun oleh

Yusuf Habibillah Akbar

18.11.2576

**PROGRAM SARJANA
PROGRAM STUDI INFORMATIKA
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2022**

**DETEKSI CLICKBAIT BERDASARKAN KOMENTAR PADA VIDEO
TRENDING DI YOUTUBE INDONESIA MENGGUNAKAN
NATURAL LANGUAGE PROCESSING**

SKRIPSI

untuk memenuhi sebagian persyaratan
mencapai gelar Sarjana
pada Program Studi Informatika



disusun oleh

Yusuf Habibillah Akbar

18.11.2576

**PROGRAM SARJANA
PROGRAM STUDI INFORMATIKA
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2022**

PERSETUJUAN

SKRIPSI

DETEKSI CLICKBAIT BERDASARKAN KOMENTAR PADA VIDEO TRENDING DI YOUTUBE INDONESIA MENGGUNAKAN NATURAL LANGUAGE PROCESSING

yang dipersiapkan dan disusun oleh

Yusuf Habibillah Akbar

13.12.6789

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 9 Agustus 2022

Dosen Pembimbing,

Ainul Yaqin, M.Kom.

NIK. 190302255

PENGESAHAN

SKRIPSI

DETEKSI CLICKBAIT BERDASARKAN KOMENTAR PADA VIDEO TRENDING DI YOUTUBE INDONESIA MENGGUNAKAN NATURAL LANGUAGE PROCESSING

yang dipersiapkan dan disusun oleh
Yusuf Habibillah Akbar

18.11.2576

telah dipertahankan di depan Dewan Penguji
pada tanggal 26 Agustus 2022

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Andi Sunyoto, M.Kom., Dr
NIK. 190302052

Ainul Yaqin, M.Kom
NIK. 190302255

Jaeni, S.Kom, M.Eng
NIK. 190302068

Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 26 Agustus 2022

DEKAN FAKULTAS ILMU KOMPUTER

Hanif Al Fatta, M.Kom
NIK. 190302096

PERNYATAAN

Saya yang bertanda tangan di bawah ini menyatakan bahwa, skripsi ini merupakan karya saya sendiri (ASLI), dan isi dalam skripsi ini tidak terdapat karya yang pernah diajukan oleh orang lain untuk memperoleh gelar akademis di suatu institusi pendidikan tinggi mana pun, dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis dan/atau diterbitkan oleh orang lain, kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan dalam daftar pustaka.

Segala sesuatu yang terkait dengan naskah dan karya yang telah dibuat adalah menjadi tanggung jawab saya pribadi.

Yogyakarta, 20 September 2022

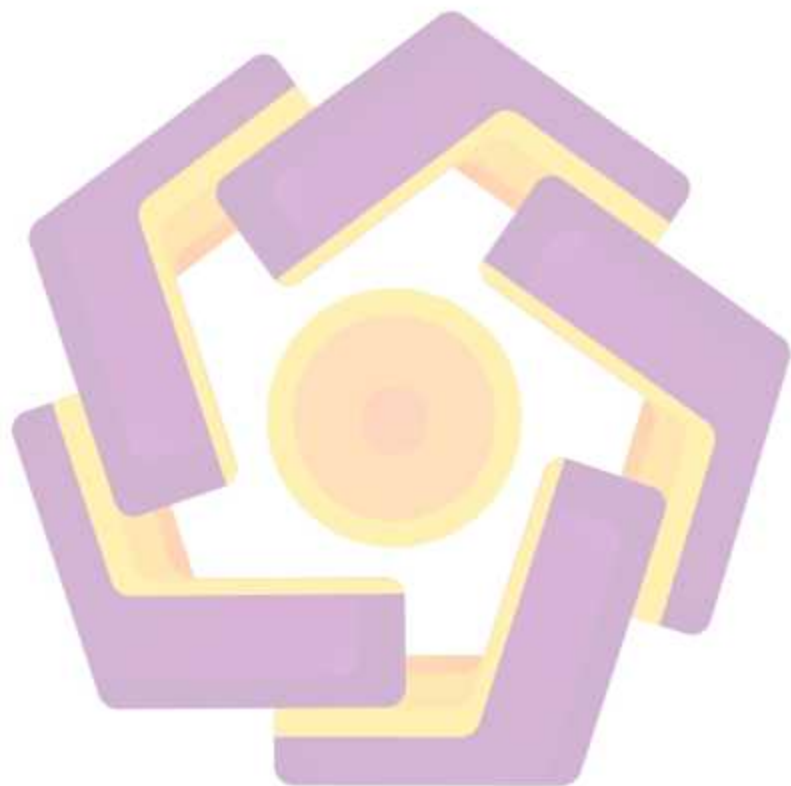


Yusuf Habibillah Akbar
NIM. 18.11.2576

MOTTO

“Ilmu adalah yang memberikan manfaat, bukan yang sekadar hanya dihafal.”

– Imam Syafi’i



PERSEMBAHAN

Alhamdulillah rabbil'alamin, puji syukur penulis panjatkan kepada Allah *Subhanahu wa ta'ala* yang telah memberikan kemampuan dan kemudahan dalam pengerjaan skripsi ini. Sholawat serta salam ter curahkan selalu kepada junjungan Nabi Agung Muhammad *Shallallahu alaihi wasallam*, yang selalu diharapkan syafaatnya di hari akhir nanti.

Bapak dan Almarhumah Ibu Tercinta sebagai tanda bakti, hormat, dan rasa terima kasih kupersembahkan karya kecil dan sederhana ini kepada Bapak dan Ibu yang telah memberikan kasih sayang, segala dukungan, dan cinta kasih yang tiada terhingga yang tidak mungkin dapat kubalas hanya dengan selembar kertas yang bertuliskan kata cinta dalam kata persembahan.

Kepada istriku yang selalu mendukung dalam setiap tahap pengerjaan skripsi, yang memberikan banyak sekali masukan dan semangat. Terima kasih telah membantu menemani berproses menyelesaikan satu-persatu tantangan dan rintangan hingga akhirnya menjadi pribadi yang mampu menuntaskan pilihannya dengan baik.

KATA PENGANTAR

Puji syukur kepada Allah SWT atas limpahan rahmat, karunia, serta petunjuk-Nya sehingga penulis dapat menyelesaikan skripsi atau penelitian dengan judul **“Deteksi *Clickbait* Berdasarkan Komentar pada Video *Trending* di YouTube Indonesia menggunakan *Natural Language Processing*”** dengan lancar. Sholawat serta salam tercurah kepada Nabi Agung Muhammad SAW, semoga kita semua mendapatkan syafaatnya di hari akhir nanti.

Penulisan skripsi ini dimaksudkan untuk memenuhi salah satu persyaratan meraih gelar sarjana strata 1 (S1), Jurusan Informatika di Fakultas Ilmu Komputer, Universitas Amikom Yogyakarta. Selain itu skripsi ini diharapkan dapat bermanfaat bagi para pembaca maupun peneliti dalam memberikan kontribusi pengetahuan khususnya dalam pendeteksian *Clickbait* pada video YouTube.

Dalam penyusunan skripsi ini penulis telah banyak mendapatkan bimbingan, arahan, bantuan, serta dukungan dari berbagai pihak. Oleh karena itu pada kesempatan ini penulis hendak mengucapkan terima kasih kepada:

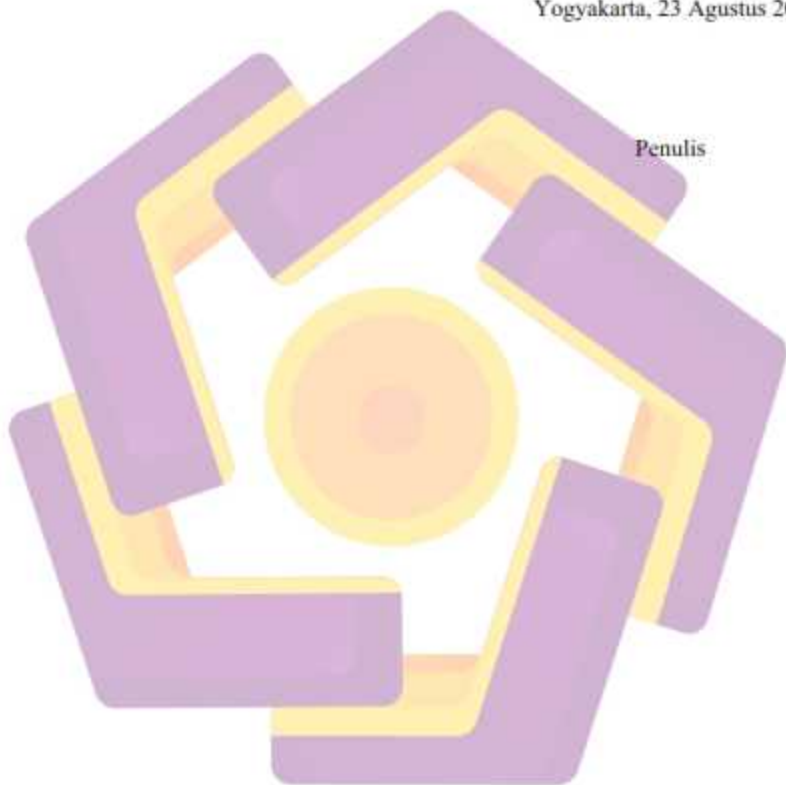
1. Bapak Hanif Al-Fatta selaku Dekan Fakultas Ilmu Komputer Universitas Amikom Yogyakarta.
2. Ibu Windha Mega Pradnya Duhita, M.Kom selaku Ketua Program Studi Informatika Fakultas Ilmu Komputer Universitas Amikom Yogyakarta.
3. Bapak Ainul Yaqin, M.Kom selaku dosen pembimbing yang selalu berbagi ilmu, arahan, dan bimbingan selama proses penyusunan skripsi ini.

4. Bapak Majid Rahardi, M.Eng, Bapak Ferian Fauzi Abdulloh, M.Kom, Bapak Kusnawi, M.Eng, dan Bapak Slamet Budiprayitno, ST., MT atas masukan serta kesempatan diskusinya selama proses penelitian.
5. Orang tua tercinta Bapak Yuli Mujiono dan Almarhumah Ibu Isnen Barokah yang telah mendidik dan senantiasa memberikan doa dan dukungan kepada saya. Kakak saya Citra Rizki Nirmala atas dukungan yang begitu besar dalam mengantarkan saya ke jenjang perkuliahan.
6. Istri tercinta Ayu Rahmawati Kautsar Dieni yang tak henti memberikan bantuan dan dukungan serta semangat kepada penulis selama penyusunan skripsi ini.
7. Keluarga besar yang senantiasa memberikan bantuan dan nasihat di saat senang maupun sulit.
8. Teman-teman di tempat kerja saya Hoshizora Foundation yang telah memberikan dukungan serta kesempatan dalam menyelesaikan studi dengan sebaik mungkin dari awal perkuliahan hingga akhir semester.
9. Seluruh teman-teman perkuliahan yang selalu memberikan dukungan moril dan bantuan kepada penulis khususnya Fajar, Rico, Wahyu, Warawandanu, Galih, Feby dan Sekar.
10. Rodhi, Azis, dan Huda yang menjadi teman bercerita dan berkeluh kesah tentang hal apa pun selama ini
11. Seluruh pihak yang membantu dalam pelaksanaan dan penyelesaian laporan skripsi serta pembelajaran dalam hidup yang mungkin tidak dapat disebutkan satu-persatu.

Akhir kata, penulis memohon maaf apabila ada kekeliruan dalam penyusunan skripsi ini. Semoga skripsi ini bermanfaat bagi penulis sendiri, masyarakat, bangsa, dan negara.

Yogyakarta, 23 Agustus 2022

Penulis

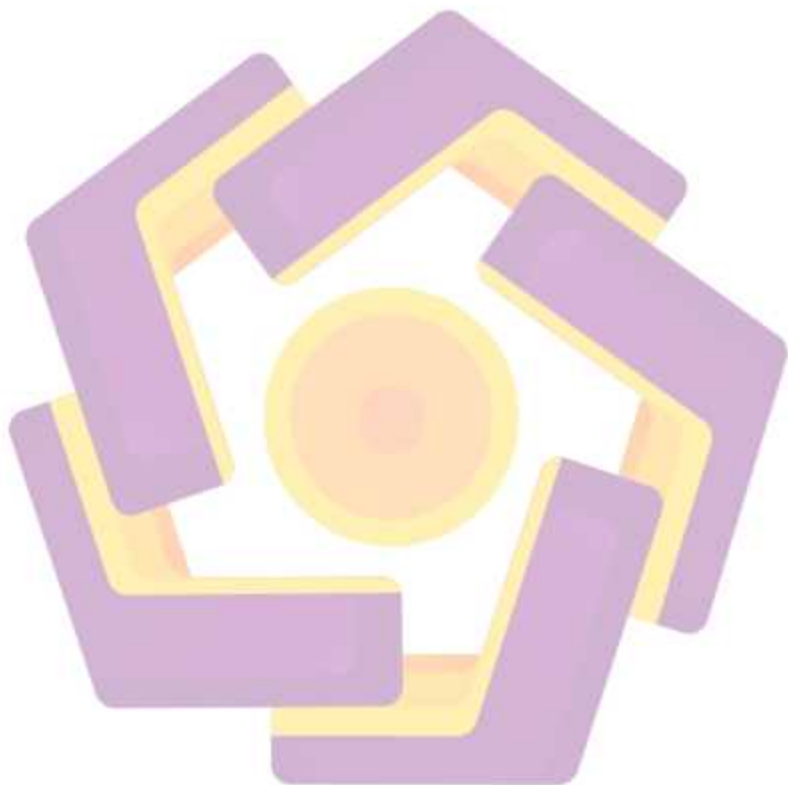


DAFTAR ISI

PERNYATAAN.....	iv
MOTTO.....	v
PERSEMBAHAN.....	vii
KATA PENGANTAR.....	viii
DAFTAR ISI.....	xi
DAFTAR TABEL.....	xiv
DAFTAR GAMBAR.....	xv
DAFTAR LAMPIRAN.....	xvii
INTISARI.....	xviii
<i>ABSTRACT</i>	xix
BAB I PENDAHULUAN	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	2
1.3 Batasan Masalah.....	3
1.4 Maksud dan Tujuan Penelitian.....	3
1.5 Manfaat Penelitian.....	3
1.6 Metode Penelitian.....	4
1.6.1 Studi Literatur.....	4
1.6.2 Identifikasi Masalah.....	4
1.6.3 Pengumpulan Data.....	4
1.6.4 Perancangan dan Implementasi.....	5
1.6.5 Pengujian.....	5
1.7 Sistematika Penulisan.....	5
BAB II LANDASAN TEORI	7
2.1 Kajian Pustaka.....	7
2.2 Dasar Teori.....	10
2.2.1 Natural Language Processing.....	10
2.2.2 <i>Sentiment Analysis</i>	10
2.2.3 <i>Naive Bayes Classifier</i>	11
2.2.4 <i>K-Nearest Neighbor</i>	11
2.2.5 <i>Cosine Similarity</i>	12

2.2.6	<i>Data Preprocessing</i>	12
2.2.7	<i>Clickbait</i>	12
BAB III METODE PENELITIAN		14
3.1	Alat dan Bahan Penelitian	14
3.1.1	Alat Penelitian	14
3.1.2	Bahan Penelitian.....	14
3.2	Alur Penelitian.....	15
3.2.1	Penyiapan Bahan	15
3.2.2	Identifikasi Masalah	19
3.2.3	Studi Literatur.....	19
3.2.4	<i>Data Selection</i>	19
3.2.5	<i>Data Annotation</i>	20
3.2.6	<i>Preprocessing</i>	21
3.2.7	Ekstraksi Fitur	23
3.2.8	<i>Modelling & Classification</i>	28
BAB IV HASIL DAN PEMBAHASAN		30
4.1	Rancangan Sistem	30
4.2	Penyiapan Data.....	30
4.2.1	<i>Data Collecting</i>	30
4.2.2	<i>Data Selection</i>	34
4.2.3	<i>Data Annotation</i>	35
4.3	Analisis Data	36
4.3.1	<i>Import Library</i>	36
4.3.2	<i>Preprocessing</i>	36
4.3.3	Feature Extraction	40
4.3.4	Data Labelling	46
4.4	<i>Modelling & Classification</i>	48
4.4.1	Train/Test Split.....	48
4.4.2	<i>Feature Scaling</i>	48
4.4.3	<i>Classification</i>	48
4.4.4	Testing	49
4.5	<i>Data Validation</i>	51

BAB V PENUTUP	52
5.1 Kesimpulan.....	52
5.2 Saran	52
Lampiran A Perbandingan Hasil Label Manual dan Prediksi	1



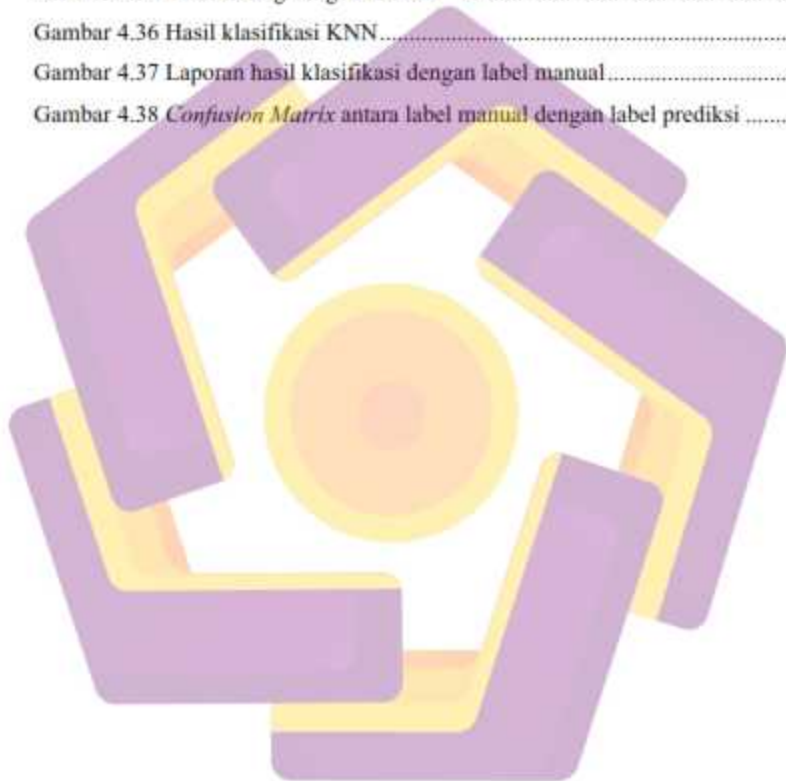
DAFTAR TABEL

Tabel 2.1 Penelitian terdahulu	8
Tabel 3.1 Daftar 10 <i>Channel</i> teratas di Indonesia	16
Tabel 3.2 Properti Dataset Video	17
Tabel 3.3 Properti Dataset Komentar Video	18
Tabel 3.4 Hasil Selection Dataset Video	20
Tabel 3.5 Hasil <i>Data Selection</i> Komentar	20
Tabel 3.6 Gambaran hasil cleaning	22
Tabel 3.7 Gambaran hasil proses formalisasi	22
Tabel 3.8 Gambaran hasil <i>stemming</i>	23
Tabel 3.9 Contoh hasil analisis sentimen	24
Tabel 3.10 Contoh hasil proses analisis <i>cosine similarity</i>	25
Tabel 3.11 Contoh hasil proses penghitungan kata	25
Tabel 3.12 Daftar <i>feature</i> dataset baru	27
Tabel 4.1 Daftar Channel Teratas Indonesia	31
Tabel 4.2 Hasil Penilaian	47

DAFTAR GAMBAR

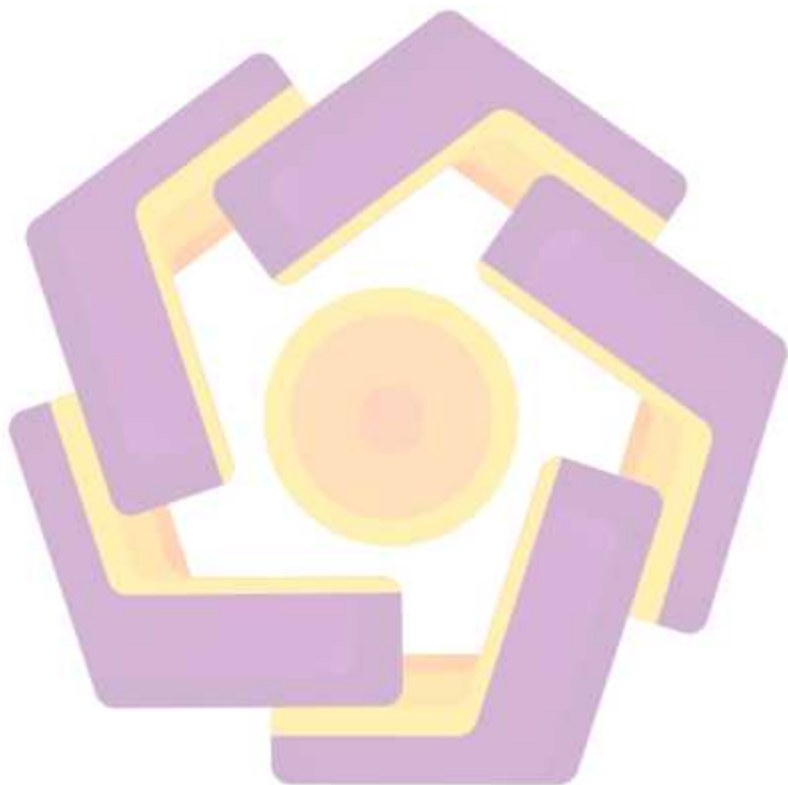
Gambar 3.1 Alur Penelitian.....	15
Gambar 3.2 Tampilan situs Social Blade	16
Gambar 3.3 Tahapan <i>Preprocessing</i>	21
Gambar 4.1 Rancangan Skema CVDI dengan Data Komentar.....	30
Gambar 4.2 Fungsi pengambil <i>metadata</i> video.....	32
Gambar 4.3 Hasil pengambilan daftar video terpopuler	32
Gambar 4.4 Fungsi pengambil <i>metadata</i> video	33
Gambar 4.5 Fungsi pengambil komentar video	33
Gambar 4.6 <i>Dataset</i> <i>metadata</i> video	34
Gambar 4.7 <i>Dataset</i> komentar video.....	34
Gambar 4.8 Hasil <i>data selection dataset metadata</i>	35
Gambar 4.9 Hasil <i>data annotation</i>	35
Gambar 4.10 Proses import <i>library</i>	36
Gambar 4.11 Fungsi tokenisasi	37
Gambar 4.12 Fungsi menghapus <i>emoji</i>	37
Gambar 4.13 Fungsi <i>cleaning data</i>	38
Gambar 4.14 Fungsi pembakuan kata	38
Gambar 4.15 Fungsi <i>stemming</i>	38
Gambar 4.16 Fungsi <i>preprocessing</i>	39
Gambar 4.17 Eksekusi fungsi <i>preprocessing</i>	39
Gambar 4.18 Hasil <i>preprocessing data komentar</i>	40
Gambar 4.19 Fungsi analisis sentimen menggunakan <i>TextBlob</i>	41
Gambar 4.20 Fungsi analisis sentimen menggunakan <i>TextBlob</i>	41
Gambar 4.21 Hasil analisis sentimen	42
Gambar 4.22 Definisi fungsi <i>cosine similarity</i>	42
Gambar 4.23 <i>Preprocessing judul</i>	43
Gambar 4.24 Fungsi analisis <i>cosine similarity</i>	44
Gambar 4.25 Hasil analisis <i>cosine similarity</i>	44
Gambar 4.26 Definisi fungsi penghitung jumlah kata.....	45
Gambar 4.27 <i>Dataset</i> komentar akhir	45
Gambar 4.28 Implementasi ekstraksi fitur	46

Gambar 4.29 Implementasi pelabelan	46
Gambar 4.30 Implementasi pembagian data <i>training & testing</i>	48
Gambar 4.31 Implementasi <i>feature scaling</i>	48
Gambar 4.32 Proses training dengan KNN Classifier.....	49
Gambar 4.33 Hasil performa klasifikasi.....	49
Gambar 4.34 Uji iterasi nilai K terhadap tingkat eror	50
Gambar 4.35 Hasil testing dengan nilai K - 13	50
Gambar 4.36 Hasil klasifikasi KNN.....	51
Gambar 4.37 Laporan hasil klasifikasi dengan label manual.....	51
Gambar 4.38 <i>Confusion Matrix</i> antara label manual dengan label prediksi	51



DAFTAR LAMPIRAN

Lampiran A Perbandingan Hasil Label Manual dan Prediksi	1
---	---



INTISARI

Clickbait merupakan salah satu bentuk strategi yang dilakukan oleh pembuat konten di YouTube untuk mendapatkan atensi penonton dengan cara membuat judul yang bombastis serta sensasional terhadap emosi pembaca. YouTube sebagai situs yang sangat populer memiliki banyak pengguna yang aktif berinteraksi baik dengan berkomentar, meninggalkan *like*, dan berbagi video ke media lainnya.

Data komentar hasil interaksi para penonton mampu menghasilkan pengetahuan baru melalui proses *Natural Language Processing* (NLP). Data komentar dapat dipelajari dan dimanfaatkan untuk mendeteksi apakah suatu video merupakan video *clickbait*.

Penelitian ini bertujuan untuk mengembangkan skema yang bernama *Clickbait Video Detection Indonesia using Comments* (CVDI-Comments) yang menggunakan data komentar untuk dianalisis apakah suatu video merupakan *clickbait* atau bukan. Hasil pengujian menunjukkan bahwa skema CVDI-Comments cukup efektif untuk mengidentifikasi video *clickbait* dibandingkan dengan identifikasi manual.

Kata Kunci: *Clickbait Detection, YouTube, Natural Language Processing, Supervised Learning*



ABSTRACT

Clickbait is one of the strategies used by content creators on YouTube to get the audience's attention by creating a bombastic and sensational title for the reader's emotions. YouTube as a very popular site has many active users who interact both by commenting, leaving likes, and sharing videos to other media.

Commentary data resulting from the interaction of the audience can generate new knowledge through the Natural Language Processing (NLP) process. Comment data can be studied and used to detect whether a video is a Clickbait video.

This study aims to develop a scheme called Clickbait Video Detection Indonesia using Comments (CVDI-Comments) which uses comment data to analyze whether a video is clickbait or not. The test results show that the CVDI-Comments scheme is quite effective for identifying clickbait videos compared to manual identification.

Keywords: *Clickbait Detection, YouTube, Natural Language Processing, Supervised Learning*



