

**ANALISIS SENTIMEN TERHADAP KOMENTAR VIDEO TRENDING
YOUTUBE MENGGUNAKAN METODE NAÏVE BAYES CLASSIFIER**

SKRIPS



disusun oleh

Youse Nur Fauzi

17.11.1219

**PROGRAM SARJANA
PROGRAM STUDI INFORMATIKA
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2021**

**ANALISIS SENTIMEN TERHADAP KOMENTAR VIDEO
TRENDING YOUTUBE MENGGUNAKAN METODE NAÏVE BAYES
CLASSIFIER**

SKRIPSI

untuk memenuhi sebagian persyaratan
mencapai gelar Sarjana
pada Program Studi Informatika



disusun oleh
Youse Nur Fauzi
17.11.1219

**PROGRAM SARJANA
PROGRAM STUDI INFORMATIKA
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2021**

PERSETUJUAN

SKRIPSI

**ANALISIS SENTIMEN TERHADAP KOMENTAR VIDEO TRENDING
YOUTUBE MENGGUNAKAN METODE NAÏVE BAYES
CLASSIFIER**

yang dipersiapkan dan disusun oleh

Youse Nur Fauzi

17.11.1219

telah disetujui oleh Dosen Pembimbing Skripsi

pada tanggal 09 Oktober 2020

Dosen Pembimbing,

Rumini, M. Kom

NIK. 190302246

PENGESAHAN

SKRIPSI

**ANALISIS SENTIMEN TERHADAP KOMENTAR VIDEO TRENDING
YOUTUBE MENGGUNAKAN METODE NAÏVE BAYES
CLASSIFIER**

yang dipersiapkan dan disusun oleh

Youse Nur Fauzi

17.11.1219

telah dipertahankan di depan Dewan Penguji
pada tanggal 21 Juli 2021

Susunan Dewan Penguji

Nama Penguji

Bety Wulan Sari, M.Kom
NIK. 190302254

Robert Marco, M.T.
NIK. 190302228

Rumini, M.Kom
NIK. 190302246

Tanda Tangan

Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 21 Juli 2021

**DEKAN FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA**

Hanif Al Fatta, M.Kom
NIK. 190302096

PERNYATAAN

Saya yang bertandatangan dibawah ini menyatakan bahwa, skripsi ini merupakan karya saya (ASLI), dan isi dalam skripsi ini tidak terdapat karya yang pernah diajukan oleh orang lain untuk memperoleh gelar akademis di suatu institusi pendidikan tinggi manapun, 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 tanggungjawab saya pribadi.

Yogyakarta, 20 September 2021

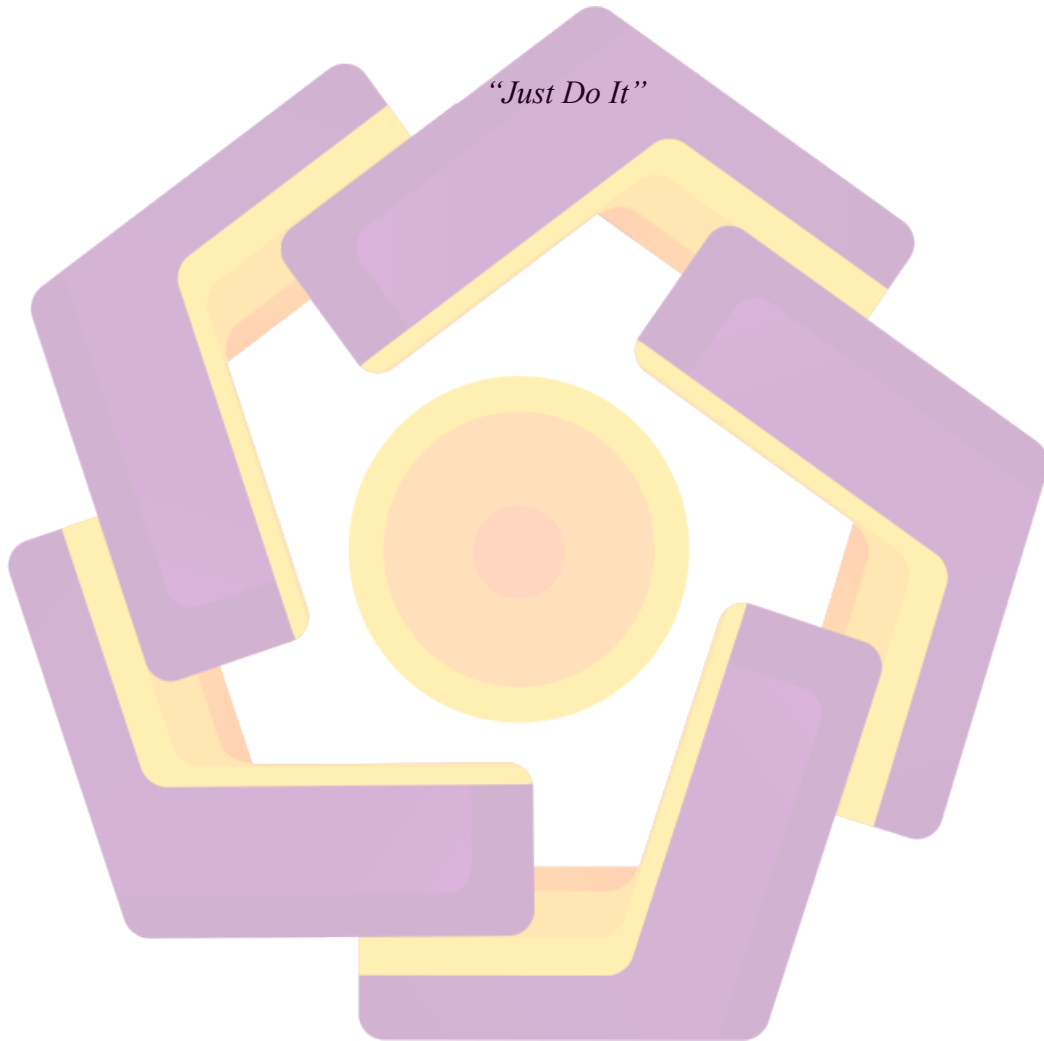


Youse Nur Fauzi

NIM. 17.11.1219

MOTTO

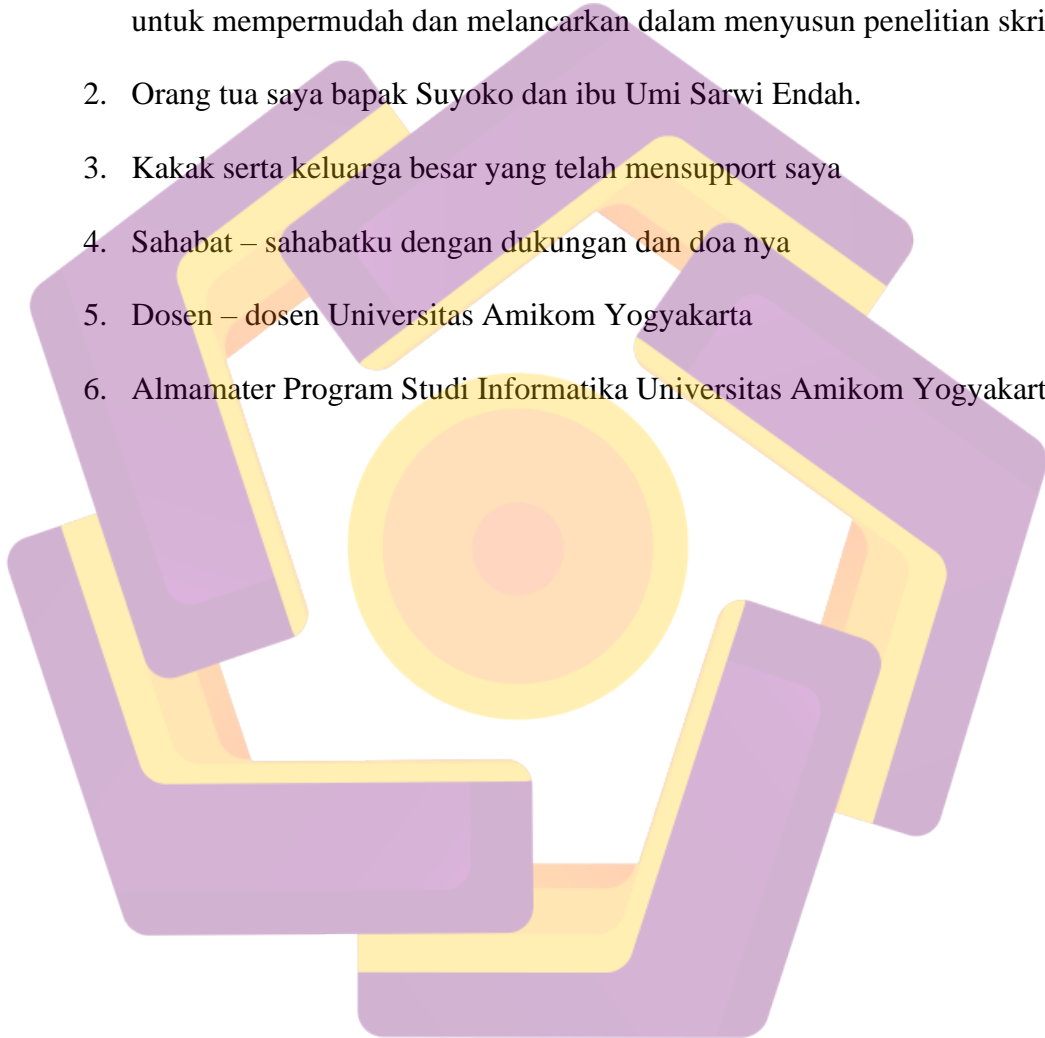
“Untuk meraih kesuksesan harus ditanamkan rasa percaya diri dan optimis untuk menjalani sesuatu yang sudah kita ambil untuk di jalani, karena tidak ada kata terlambat untuk bangkit dari keterpurukan atas keputusan yang sudah kita ambil”



PERSEMBAHAN

Skripsi ini saya persembahkan untuk:

1. Allah SWT yang telah senantiasa memberikan rahmat dan hidayah-Nya untuk mempermudah dan melancarkan dalam menyusun penelitian skripsi.
2. Orang tua saya bapak Suyoko dan ibu Umi Sarwi Endah.
3. Kakak serta keluarga besar yang telah mensupport saya
4. Sahabat – sahabatku dengan dukungan dan doanya
5. Dosen – dosen Universitas Amikom Yogyakarta
6. Almamater Program Studi Informatika Universitas Amikom Yogyakarta



KATA PENGANTAR

Puji dan syukur kami panjatkan kehadirat Allah SWT, karena hanya atas berkat dan rahmat-Nya, sehingga Laporan Tugas Akhir yang berjudul **“ANALISIS SENTIMEN TERHADAP KOMENTAR VIDEO TRENDING YOUTUBE MENGGUNAKAN METODE NAÏVE BAYES CLASSIFIER”** dapat selesai. Tugas Akhir ini merupakan syarat untuk menyelesaikan program studi Strata Satu di Fakultas Ilmu Komputer pada Universitas Amikom Yogyakarta.

Selama pelaksanaan Tugas Akhir hingga selesainya laporan Tugas Akhir ini dapat terwujud dan terselesaikan berkat bimbingan dan saran dari berbagai pihak. Pada kesempatan ini, penulis ingin menyampaikan terima kasih kepada:

1. Allah SWT yang telah memberikan rahmat, nikmat, serta karunia-Nya kepada penulis.
2. Orang tua serta keluarga penulis untuk semua yang telah diberikan kepada penulis, atas perjuangan yang tiada lelah untuk memberikan yang terbaik, kasih sayang tiada terhingga serta segala doa dan restu mereka yang senantiasa menyertai dalam setiap langkah penulis.
3. Ibu Windha Mega Pradnya D, M.Kom, selaku Ketua Program Studi Informatika dan Bapak Arif Akbarul Huda, S.Si, M.Eng, selaku Sekretaris Program Studi Informatika.
4. Ibu Rumini, M.Kom, selaku Dosen Pembimbing yang telah memberikan bimbingan, masukan, serta saran sehingga skripsi ini dapat terselesaikan dengan baik.

5. Bapak dan Ibu Dosen Program Studi Informatika Fakultas Ilmu Komputer Universitas Amikom Yogyakarta yang telah memberikan ilmu yang bermanfaat kepada penulis baik didalam maupun diluar kelas.
6. Teman-teman KOSAN yang telah membantu, menemani dan mendukung dalam menyelesaikan skripsi ini.
7. Teman-teman 17 Informatika 05 yang telah membantu dalam perkuliahan selama 4 tahun.
8. Serta terima kasih kepada semua pihak yang tidak dapat penulis sebutkan satu persatu yang telah membantu penulis selama ini.

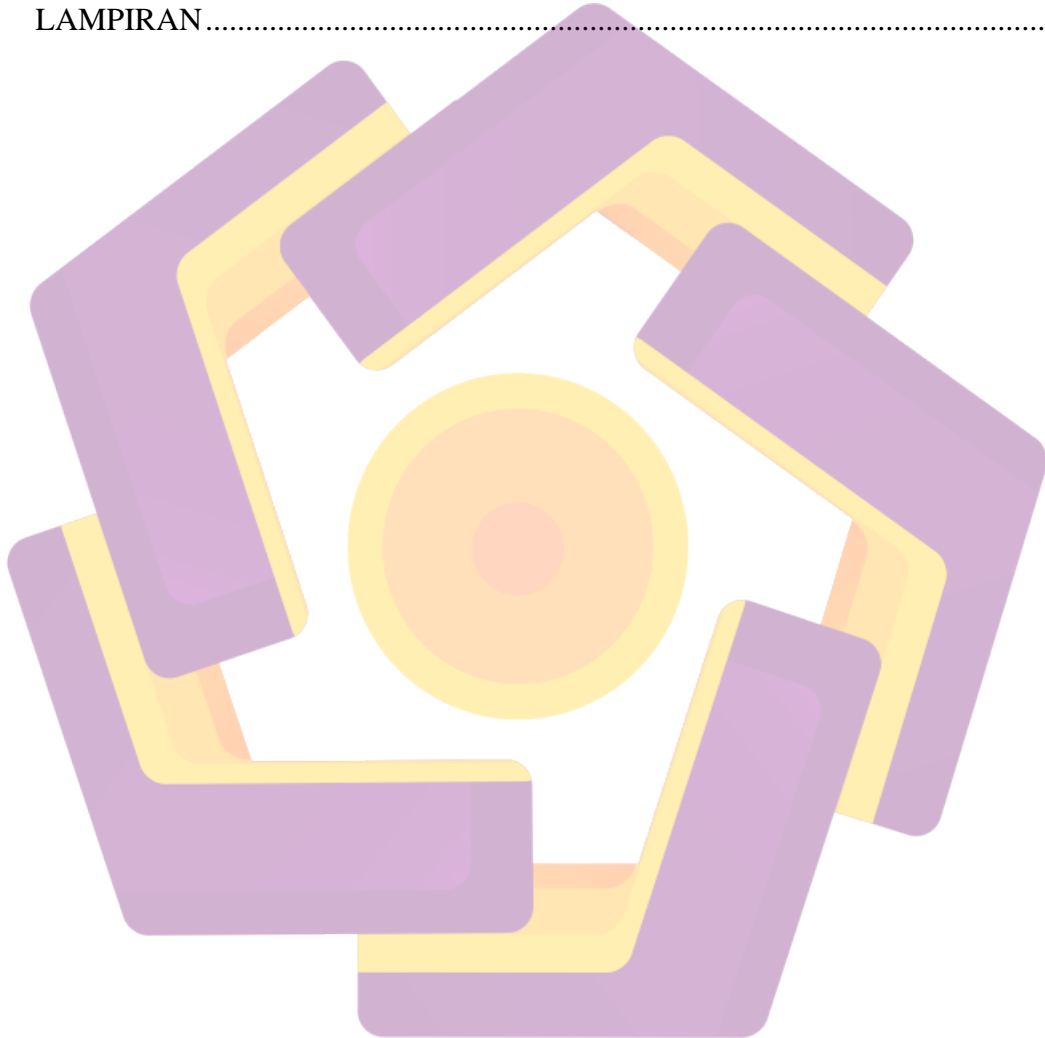
Akhir kata, penulis menyadari bahwa laporan skripsi ini masih jauh dari sempurna karena terbatasnya pengetahuan dan pengalaman yang dimiliki penulis. Oleh karena itu penulis memohon maaf atas semua kekurangan yang terdapat dalam skripsi ini, serta penulis mengharapkan segala bentuk masukan dan kritik dari berbagai pihak. Semoga skripsi ini dapat bermanfaat bagi para pembaca dan semua pihak dalam hal perancangan user interface dan analisa user experience sebuah website.

DAFTAR ISI

HALAMAN JUDUL.....	i
PERSETUJUAN	ii
PENGESAHAN	iii
PERNYATAAN.....	iv
MOTTO	v
PERSEMBAHAN.....	vi
KATA PENGANTAR	vii
DAFTAR ISI.....	ix
DAFTAR TABEL.....	xii
DAFTAR GAMBAR	xiii
INTISARI.....	xv
ABSTRACT.....	xvi
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	2
1.3 Batasan Masalah.....	2
1.4 Manfaat dan Tujuan Penelitian.....	3
1.4.1 Tujuan Penelitian	3
1.4.2 Manfaat Penelitian	3
1.5 Metode Penelitian.....	4
1.5.1 Metode Pengumpulan Data.....	4
1.5.2 Metode Analisis	5
1.5.3 Metode Pengujian.....	5
1.6 Sistematika Penulisan.....	6
BAB II LANDASAN TEORI.....	9

2.1	Tinjauan Pustaka	9
2.2	Dasar Teori	13
2.2.1	Analisis Sentimen	13
2.2.2	Data Mining	14
2.2.3	Preprocessing	15
2.2.4	Feature Selection	19
2.2.5	Klasifikasi	20
2.2.6	Metode Naïve Bayes	23
2.2.7	Confusion Matrix	25
2.2.8	Python	26
2.2.9	Visual Studio Code	27
BAB III METODE PENELITIAN.....		28
3.1	Alur Penelitian.....	28
3.2	Alat dan Bahan Penelitian	29
3.2.1	Alat.....	30
3.2.2	Bahan.....	30
3.3	Metode Analisis.....	31
3.3.1	Pengumpulan Data	31
3.3.2	Pengolahan Awal Data	33
3.3.3	Klasifikasi Naive Bayes	46
3.4	Perancangan Sistem.....	47
3.4.1	Flowchart	48
3.5	Perancangan Antarmuka Pengguna	49
3.5.1	Halaman Utama.....	49
3.5.2	Halaman Proses Prediksi Data Crawling	50
BAB IV HASIL DAN PEMBAHASAN		52
4.1	Pengujian	52
4.1.1	Confusion Matrix	52
4.1.2	K-Fold Cross Validation	55
4.1.3	Hasil Prediksi	56
4.2	Implementasi Antarmuka	56
4.2.1	Halaman Utama.....	57

4.2.2	Halaman Proses Prediksi Sentimen.....	58
BAB V PENUTUP.....		62
5.1	Kesimpulan.....	62
5.2	Saran.....	62
DAFTAR PUSTAKA		64
LAMPIRAN.....		66



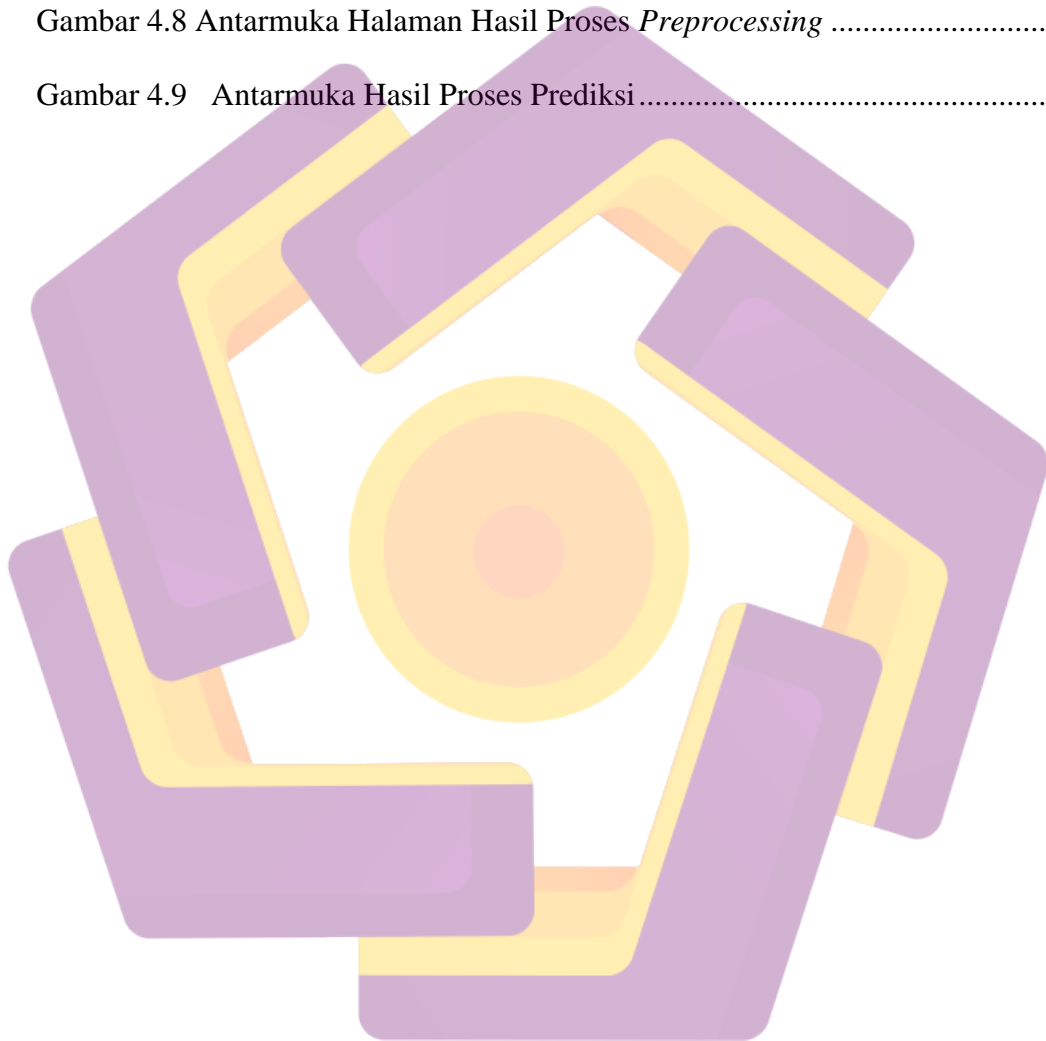
DAFTAR TABEL

Tabel 2.1 Perbandingan dengan penelitian-penelitian sebelumnya	11
Tabel 2.2 <i>Confusion Matrix</i>	25
Tabel 3.1 Hasil <i>Crawling</i> data <i>Comments</i>	33
Tabel 3.2 Hasil Data yang di Labelling Secara Manual Sebagai Dataset	33
Tabel 3.3 Hasil Data <i>Case Folding</i>	35
Tabel 3.4 Data Setelah Memperbaiki <i>Slang Word</i>	37
Tabel 3.5 Data Setelah <i>Cleansing</i>	39
Tabel 3.6 Hasil Akhir <i>Text Preprocessing</i>	42
Tabel 3.7 Hasil TF-IDF	45
Tabel 4.1 <i>Confusion Matrix Multinomial</i> – TF-IDF Pengujian Pertama	52
Tabel 4.2 <i>Confusion Matrix Multinomial</i> – TF-IDF Pengujian Kedua	53
Tabel 4.3 <i>Confusion Matrix Multinomial</i> – TF-IDF Pengujian Ketiga	54

DAFTAR GAMBAR

Gambar 2.1 Tahapan Preprocessing.....	17
Gambar 2.2 <i>Learning</i>	21
Gambar 2.3 <i>Classification</i>	22
Gambar 3.1 Diagram alur penelitian.....	28
Gambar 3.2 <i>Script Crawling Data YouTube</i>	32
Gambar 3.3 <i>Script Case Folding Data</i>	35
Gambar 3.4 <i>Script Slang Word Data</i>	37
Gambar 3.5 <i>Script Cleansing Data</i>	39
Gambar 3.6 <i>Script Pada Tahap Tokenizing, Stopword Removal dan Stemming</i> ...	42
Gambar 3.7 <i>Script Pembagian Data Latih dan Data Uji</i>	44
Gambar 3.8 <i>Script TF-IDF</i>	45
Gambar 3.9 <i>Script Model Klasifikasi</i>	46
Gambar 3.10 Klasifikasi dengan Kernel <i>Multinomial</i> dan TF-IDF	47
Gambar 3.11 <i>Script</i> untuk menyimpan model	47
Gambar 3.12 <i>Flowchart Sistem</i>	48
Gambar 3.13 Rancangan Halaman Home dan Trending Video.....	50
Gambar 3.14 Rancangan Halaman untuk Menampilkan Hasil <i>Crawling</i>	50
Gambar 3.15 Rancangan Halaman untuk Menampilkan Hasil <i>Preprocessing</i>	51
Gambar 3.16 Rancangan Halaman untuk Hasil Prediksi	51
Gambar 4.1 <i>Classification Report Polynomial</i> – TF-IDF Pengujian Pertama.....	53
Gambar 4.2 <i>Classification Report Polynomial</i> – TF-IDF Pengujian Kedua	54
Gambar 4.3 <i>Classification Report Polynomial</i> – TF-IDF Pengujian Ketiga	55

Gambar 4.4 Hasil <i>K-Fold Cross Validation</i> Kernel <i>Multinomial</i> – TF-IDF	55
Gambar 4.5 Hasil Prediksi Berbentuk Diagram Lingkaran	56
Gambar 4.6 Antarmuka Halaman Utama.....	58
Gambar 4.7 Antarmuka Halaman Proses Hasil <i>Crawling</i>	59
Gambar 4.8 Antarmuka Halaman Hasil Proses <i>Preprocessing</i>	60
Gambar 4.9 Antarmuka Hasil Proses Prediksi.....	61



INTISARI

YouTube merupakan salah satu media untuk berbagi video yang memungkinkan pengguna untuk mengunggah, menonton dan berbagi video. Di dalam YouTube terdapat salah satu menu yang menyajikan video-video yang sedang populer (trending tab). Sehingga akan dilakukan analisis sentimen untuk mengetahui kecenderungan komentar masyarakat terhadap video yang sedang trending, apakah lebih cenderung berisikan komentar positif atau komentar negatif. Dalam penelitian ini data yang digunakan yaitu komentar yang di kumpulkan dari video YouTube yang sedang trending, dengan menggunakan salah satu kernel metode Naïve Bayes yaitu *multinomial* dan dilakukan pembobotan kata pada dokumen yaitu menggunakan *Term Frequency-Inverse Document Frequency* (TF-IDF).

Untuk melakukan klasifikasi, langkah awal yang harus dilakukan yaitu *crawling* data komentar menggunakan YouTube API, data yang berhasil dikumpulkan akan diberikan label secara manual terlebih dahulu. Setekah itu akan dilakukan *preprocessing* terhadap data, data yang sudah bersih selanjutnya akan diberikan pembobotan pada kata menggunakan metode TF-IDF. Lalu langkah terakhir data akan dilakukan klasifikasi menggunakan kernel *Multinomial* pada Naïve Bayes.

Untuk pengujian performa akan menggunakan metode *confusion matrix* dan *K-Fold Cross Validation*.

Kata Kunci: Kernel Multinomial NB, TF-IDF, Analisis sentimen

ABSTRACT

YouTube is one of the media for sharing videos that allows users to upload, watch, and share videos. In YouTube there is a menu that presents videos that are currently popular (trending tab). So that a sentiment analysis will be carried out to determine the public's tendency towards trending videos, whether they tend to contain positive comments or negative comments. In this study, the data used are comments collected from trending YouTube videos, using one of the Naïve Bayes kernel methods, namely multinomial and carrying out word-weighting on documents using Term Frequency-Inverse Document Frequency (TF-IDF).

To do the classification, the first step that must be done is crawling comment data using the YouTube API, the data that has been collected will be manually labeled first. After that, preprocessing will be carried out on the data, the clean data will then be given a weighting on the words using the TF-IDF method. Then the last step is to classify the data using the Multinomial kernel in Nave Bayes.

For the performance of the test, the confusion matrix and K-Fold Cross Validation will be used.

Keyword: *NB Multinomial Kernel, TF-IDF, Sentiment Analysis*

