

**SENTIMEN ANALISIS MUSIK K-POP BERDASARKAN YOUTUBE  
DENGAN METODE KLASIFIKASI NAÏVE BAYES**

**TUGAS AKHIR**



disusun oleh

**Moch. Alf Dwi Mahendrawan**

**17.11.0976**

**PROGRAM SARJANA  
PROGRAM STUDI TEKNIK INFORTIKA  
FAKULTAS ILMU KOMPUTER  
UNIVERSITAS AMIKOM YOGYAKARTA  
YOGYAKARTA  
2021**

**SENTIMEN ANALISIS MUSIK K-POP BERDASARKAN YOUTUBE  
DENGAN METODE KLASIFIKASI NAÏVE BAYES**

**TUGAS AKHIR**

untuk memenuhi sebagian persyaratan mencapai gelar Sarjana Komputer pada  
jenjang Program Sarjana – Program Studi Teknik Informatika



disusun oleh

**Moch. Alif Dwi Mahendrawan**

**17.11.0976**

**PROGRAM SARJANA  
PROGRAM STUDI TEKNIK INFORTIKA  
FAKULTAS ILMU KOMPUTER  
UNIVERSITAS AMIKOM YOGYAKARTA  
YOGYAKARTA  
2021**

**PERSETUJUAN**

**TUGAS AKHIR**

**SENTIMEN ANALISIS MUSIK K-POP BERDASARKAN YOUTUBE  
DENGAN METODE KLASIFIKASI NAÏVE BAYES**

yang dipersiapkan dan disusun oleh

**Moch. Alif Dwi Mahendrawan**

**17.11.0976**

telah disetujui oleh Dosen Pembimbing Skripsi  
pada tanggal 7 Juli 2021

**Dosen Pembimbing,**

**Banu Santoso, S.T., M.Eng.**  
**NIK. 190302327**

## PENGESAHAN

## TUGAS AKHIR

### SENTIMEN ANALISIS MUSIK K-POP BERDASARKAN YOUTUBE DENGAN METODE KLASIFIKASI NAÏVE BAYES

yang dipersiapkan dan disusun oleh

**Moch. Alf Dwi Mahendrawan**  
17.11.0976

telah dipertahankan di depan Dewan Penguji  
pada tanggal 19 Juli 2021

**Susunan Dewan Penguji**

**Nama Penguji**

**Tanda Tangan**

M.Rudyanto Arief, S.T., M.T  
NIK. 190302098

Andriyan Dwi Putra M.Kom  
NIK. 190302270

Banu Santoso, S.T., M.Eng  
NIK. 190302327

Tugas Akhir ini telah diterima sebagai salah satu persyaratan  
untuk memperoleh gelar Sarjana Komputer  
Tanggal 19 Juli 2021

**DEKAN FAKULTAS ILMU KOMPUTER**

Hanif Al Fatta,S.Kom., M.Kom.  
NIK. 190302096

## PERNYATAAN

Saya yang bertandatangan dibawah 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 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, 02-08-2021



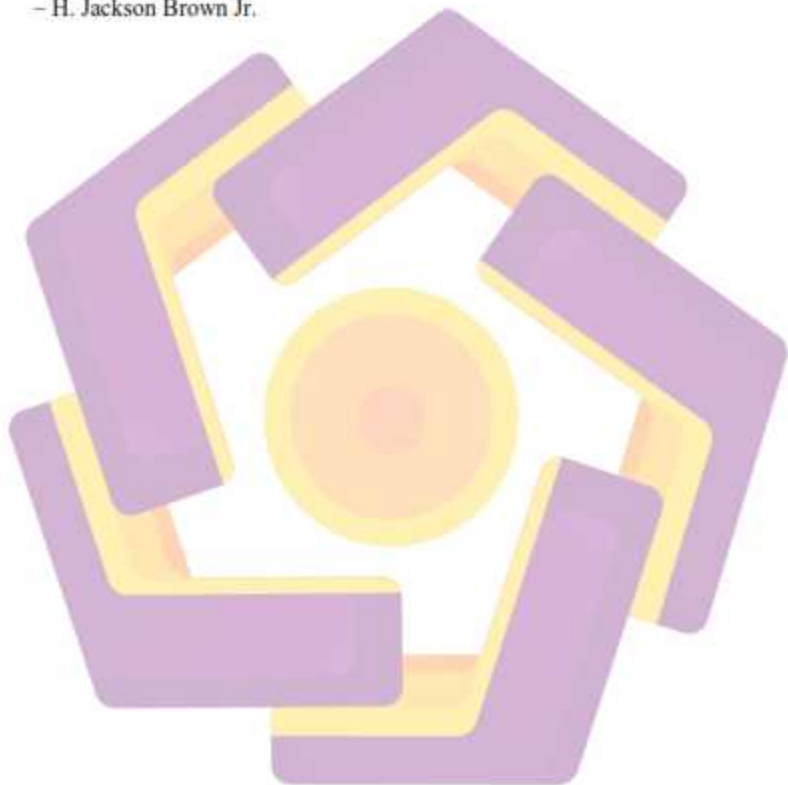
Moch Alif Dwi Mahendrawan

NIM. 17.11.0975

## MOTTO

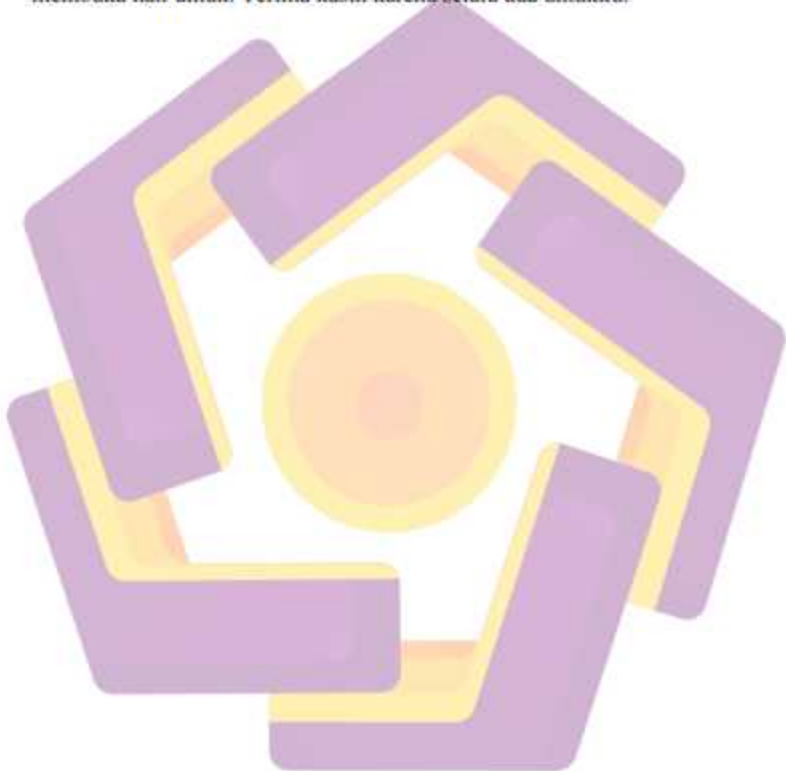
“Don’t say you don’t have enough time. You have exactly the same number of hours per day that were given to Helen Keller, Pasteur, Michelangelo, Mother Teresa, Leonardo da Vinci, Thomas Jefferson, and Albert Einstein.”

– H. Jackson Brown Jr.



## PERSEMBAHAN

Skripsi ini adalah persembahan kecil saya untuk kedua orangtua saya. Ketika dunia menutup pintunya pada saya, ayah dan ibu membuka lengannya untuk saya. Ketika orang-orang menutup telinga mereka untuk saya, mereka berdua membuka hati untuk. Terima kasih karena selalu ada untukku.



## KATA PENGANTAR

Segala puji bagi Allah SWT yang telah memberikan rahmat dan karunia Nya kepada penulis, sehingga penulis dapat menyelesaikan skripsi ini dengan baik. Shalawat dan salam senantiasa tercurah kepada Rasulullah SAW yang mengantarkan manusia dari zaman kegelapan ke zaman yang terang benderang ini. Penyusunan skripsi ini dimaksudkan untuk memenuhi sebagian syarat-syarat guna mencapai gelar Sarjana Komputer di Universitas Amikom Yogyakarta.

Penulis menyadari bahwa penulisan ini tidak dapat terselesaikan tanpa dukungan dari berbagai pihak baik moril maupun materil. Oleh karena itu, penulis ingin menyampaikan ucapan terima kasih kepada semua pihak yang telah membantu dalam penyusunan skripsi ini terutama kepada:

1. Kedua orang tua saya yang telah memberikan dukungan baik moril maupun materil serta doa yang tiada henti-hentinya kepada saya.
2. Segenap keluarga dan teman yang telah menyemangati dan membantu penyelesaian skripsi ini.
3. Bapak Banu Santoso, S.T., M.Eng. selaku dosen Pembimbing Skripsi saya yang telah bersedia membimbing dan mengarahkan penulis selama menyusun skripsi dan memberikan banyak ilmu serta solusi pada setiap permasalahan atas kesulitan dalam penulisan skripsi ini.
4. Seluruh Bapak/Ibu dosen Fakultas ilmu komputer yang telah memberikan pengetahuan yang sangat bermanfaat selama masa perkuliahan.



5. Seluruh teman-teman seangkatan, terutama kelas Informatika 1 Angkatan 2017 yang selalu mengisi hari-hari menjadi sangat menyenangkan.
6. Seluruh staf dan karyawan Universitas Amikom Yogyakarta yang telah memberikan bantuan kepada penulis.
7. Sahabat saya Nur Rahman dan kakak saya Moch Herlambang, yang telah membantu memberikan solusi dan ide dalam penyelesaian skripsi ini.

Penulis menyadari bahwa skripsi ini masih jauh dari sempurna dikarenakan terbatasnya pengalaman dan pengetahuan yang dimiliki penulis. Oleh karena itu, penulis mengharapkan segala bentuk saran serta masukan bahkan kritik yang membangun dari berbagai pihak. Semoga skripsi ini dapat bermanfaat bagi para pembaca dan semua pihak khususnya dalam bidang teknik informatika.

Yogyakarta, 2 Agustus 2021

Penulis,



(Moch. Alif Dwi Mahendrawan)

# DAFTAR ISI

SENTIMEN ANALISIS MUSIK K-POP BERDASARKAN YOUTUBE DENGAN METODE KLASIFIKASI NAIVE BAYES .....	I
PERNYATAAN .....	IV
MOTTO .....	VI
PERSEMBAHAN .....	VII
KATA PENGANTAR .....	VIII
DAFTAR ISI .....	X
DAFTAR TABEL .....	XII
DAFTAR GAMBAR .....	XIII
DAFTAR ISTILAH .....	XV
INTISARI .....	XVI
ABSTRACT .....	XVII
<b>BAB I PENDAHULUAN .....</b>	<b>1</b>
1.1 LATAR BELAKANG .....	1
1.2 RUMUSAN MASALAH .....	2
1.3 BATASAN MASALAH .....	3
1.4 MAKSUD DAN TUJUAN PENELITIAN .....	4
1.5 MANFAAT PENELITIAN .....	5
1.6 METODE PENELITIAN .....	5
1.6.1 Metode Pengumpulan Data .....	5
1.6.2 Metode Analisis .....	5
1.6.3 Metode Perancangan .....	6
1.7 SISTEMATIKA PENULISAN .....	7
<b>BAB II LANDASAN TEORI .....</b>	<b>8</b>
2.1 KAJIAN PUSTAKA .....	8
2.2 DASAR TEORI .....	14
2.2.1 YouTube .....	14
2.2.2 Text Mining .....	14

2.2.3	<i>Analisis Sentimen</i>	15
2.2.4	<i>Scraping</i>	16
2.2.5	<i>Text Pre-Processing</i>	16
2.2.6	<i>Term Frequency</i>	18
2.2.7	<i>Naïve Bayes Classifier</i>	19
2.2.8	<i>Confusion Matrix</i>	21
<b>BAB III METODE PENELITIAN</b>		<b>23</b>
3.1	<b>ALAT DAN BAHAN PENELITIAN</b>	23
3.2	<b>ALUR PENELITIAN</b>	23
4.2.1	<i>Ishikawa Diagram</i>	24
4.2.2	<i>General Arsitektur Sistem</i>	25
<b>BAB IV HASIL DAN PEMBAHASAN</b>		<b>41</b>
4.1	<b>RANCANGAN SISTEM</b>	41
4.1.1	<i>Perancangan Perangkat Lunak</i>	41
4.2	<b>ALUR PRODUKSI</b>	64
4.2.1	<i>Use Case Diagram</i>	64
4.2.2	<i>Activity Diagram</i>	65
4.2.3	<i>Class Diagram</i>	68
4.3	<b>PEMBUATAN PRODUK</b>	69
4.4	<b>HASIL AKHIR PRODUK</b>	72
4.5	<b>HASIL PENGUJIAN DAN PEMBAHASAN</b>	75
<b>BAB V PENUTUP</b>		<b>80</b>
5.1	<b>KESIMPULAN</b>	80
5.2	<b>SARAN</b>	80
<b>DAFTAR PUSTAKA</b>		<b>82</b>

## DAFTAR TABEL

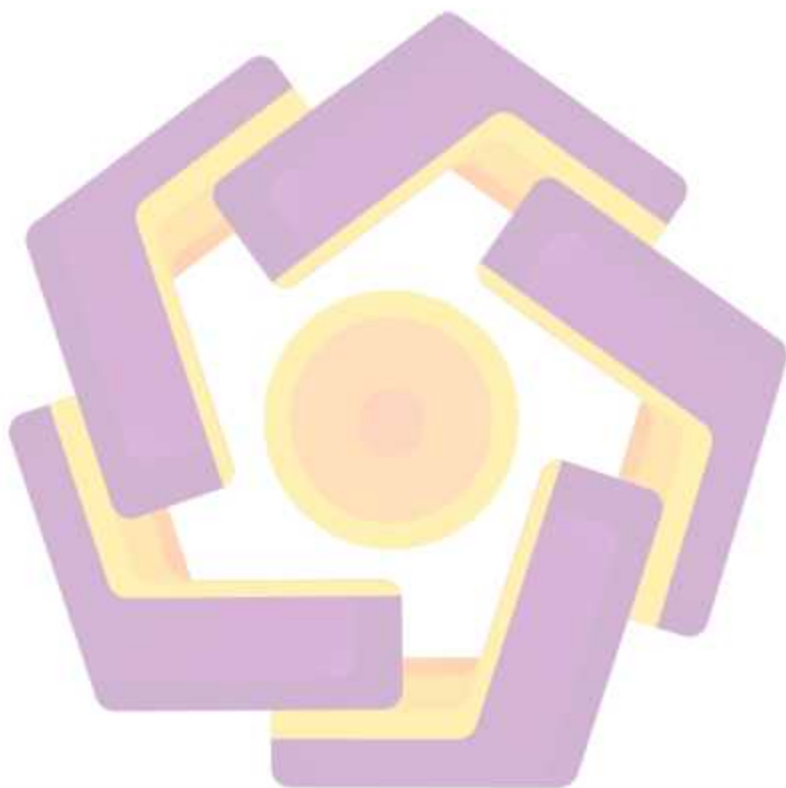
Tabel 2.1	Ikhtisar Pustaka .....	10
Tabel 2.2	Contoh tahap Case folding pada teks .....	17
Tabel 2.3	Contoh tahap Filtering pada teks .....	18
Tabel 2.4	Contoh tahap Stemming pada teks .....	19
Tabel 2.5	Perbandingan Naïve Bayes dengan SVM dan logistic regression...	21
Tabel 2.6	Confusion Matrix.....	22
Tabel 3.1	Data Kalimat.....	31
Tabel 3.2	Kumpulan Kata Pada Kalimat .....	31
Tabel 3.3	Penyusunan Daftar Kata dan Perhitungan .....	32
Tabel 3.4	Perhitungan Nilai IDF .....	33
Tabel 3.5	Perhitungan Nilai TF-IDF .....	34
Tabel 3.6	Contoh Kalimat Klasifikasi .....	35
Tabel 4.1	Atribut Preprocessing .....	40
Tabel 4.2	Prosedur Proses .....	41
Tabel 4.3	Prosedur File Stopword .....	43
Tabel 4.4	Atribut TF-IDF .....	44
Tabel 4.5	Prosedur Perhitungan TF .....	46
Tabel 4.6	Prosedur Perhitungan IDF .....	48
Tabel 4.7	Prosedur Perhitungan TF-IDF .....	50
Tabel 4.8	Atribut Klasifikasi Naïve Bayes .....	51
Tabel 4.9	Prosedur CountPrior .....	53
Tabel 4.10	Prosedur countWordInClass .....	55
Tabel 4.11	Prosedur countVocabulary .....	57
Tabel 4.12	Prosedur countProbability .....	59
Tabel 4.13	Prosedur isClass.....	61
Tabel 4.14	Use Case Scenario Menampilkan Hasil Analisis Sentimen .....	63
Tabel 4.15	Data Komentar Lagu Korea.....	70
Tabel 4.16	Hasil Confusion Matrix DUN DUN/Everglow .....	71
Tabel 4.17	Hasil Confusion Matrix Wannabe/Itzy .....	71
Tabel 4.18	Pengujian Sistem Analisis Sentimen .....	72

## DAFTAR GAMBAR

Gambar 3.1	Ishikawa Diagram.....	24
Gambar 3.2	General Arsitektur Sistem .....	26
Gambar 3.3	Flowchart Sistem Analisis Sentimen.....	27
Gambar 3.4	Memberikan Alamat Url.....	28
Gambar 3.5	Memilih Data Konversi .....	28
Gambar 3.6	Proses Scrapping .....	28
Gambar 3.7	Konversi .txt .....	28
Gambar 3.8	Mempersiapkan File .txt.....	28
Gambar 3.9	Labelisasi DUNDUN/Everglow.....	30
Gambar 3.10	Dataset Setelah Melalui Preprocessing .....	38
Gambar 3.11	Contoh Hasil Analisis Sentimen Lagu Forever/Aespa.....	38
Gambar 4.1	Flowchart Class Preprocessing.....	41
Gambar 4.2	Flowchart file Stopword.....	42
Gambar 4.3	Flowchart TF.....	45
Gambar 4.4	Flowchart IDF .....	47
Gambar 4.5	Flowchart Perhitungan TF-IDF .....	49
Gambar 4.6	Flowchart CountPrior.....	52
Gambar 4.7	Flowchart countWordInClass.....	54
Gambar 4.8	Flowchart countVocabulary .....	56
Gambar 4.9	Flowchart countProbability .....	58
Gambar 4.10	Flowchart IsClass .....	60
Gambar 4.11	Use Case Diagram .....	62
Gambar 4.12	Input Text File Activity Diagram .....	64
Gambar 4.13	Show Result Data Activity Diagram .....	65
Gambar 4.14	Class Diagram Analisis Sentimen .....	66
Gambar 4.15	Flowchart Program Aplikasi Analisis Sentimen .....	67
Gambar 4.16	Menu Utama .....	67
Gambar 4.17	Rating .....	68
Gambar 4.18	Analisa.....	69
Gambar 4.19	Browse.....	69



Gambar 4.20	Result.....	70
Gambar 4.21	Hasil Analisis Lagu DUN DUN/Everglow .....	71
Gambar 4.21	Hasil Analisis Lagu Wannabe/Itzy.....	71



## DAFTAR ISTILAH

- *Channel* Sebuah saluran yang ada dalam aplikasi YouTube
- *Subscribe* Sebuah tombol untuk ber langganan ke saluran yang kita inginkan
- *Like* Sebuah tombol yang memiliki arti kita menyukai video tersebut
- *Trending* Sebuah topik yang lagi dicari oleh banyak orang dalam hal ini berarti melalui YouTube
- *Supervised-learning* adalah operasi *machine learning* yang sering digunakan kepada data yang mana ada sebuah pemetaan yang benar diantara data input dan output.
- *Talented* adalah bertalenta atau berbakat
- *Naïve bayes classifier* atau merupakan klasifikasi naïve bayes yang berdasarkan rumus teorema bayes
- *scrapping* berarti mengumpulkan atau menggali dalam hal ini data

## INTISARI

Dunia musik tanah air sekarang tengah kebanjiran musik-musik luar negeri yang kini semakin digandrungi oleh para remaja. terbukti dari semakin banyak bermunculan lagu-lagu k-pop yang sering *trending* di Indonesia. Jumlah musik k-pop yang banyak tentunya memiliki jumlah pendengar yang lebih banyak lagi. Namun sayangnya, dengan banyaknya lagu baru ciptaan *group band* kita sukai banyak yang tidak diimbangi dengan kualitas lagunya. Banyak lagu k-pop yang lebih mengedepankan aspek hiburan tanpa memikirkan kualitas lagunya.

Analisis sentimen atau *text mining* adalah studi komputasi opini publik, sentimen dan emosi melalui entitas dan atribut yang biasa kita ungkapkan dalam bentuk teks. Analisis *sentimen* akan mengelompokkan polaritas teks yang ada dalam kalimat atau *dokumen* untuk mengetahui opini yang diungkapkan dalam kalimat atau *dokumen* tersebut berupa *positif* atau *negatif*.

Penelitian ini menghasilkan rekomendasi musik k-pop dari komentar *positif* dan *negatif* di YouTube *trending* menggunakan metode Naïve Bayes Classifier. Hasil dari akurasi klasifikasi *naïve bayes* adalah 79% setelah diuji dengan menggunakan metode *confusion matrix*.

**Kata Kunci:** Analisis sentimen, Naïve Bayes Classifier, k-pop, Komentar *negatif*, Komentar *positif*, YouTube



## ABSTRACT

The world of national music is now being flooded with foreign music which is now increasingly loved by teenagers. as evidenced by the increasing number of k-pop songs that are often trending in Indonesia. The large number of k-pop music certainly has an even greater number of listeners. But unfortunately, with the number of new songs created by their bands, many of which are not matched by the quality of the songs. Many k-pop songs prioritize the entertainment aspect without thinking about the quality of the song.

Sentiment analysis or text mining is the computational study of public opinion, sentiment and emotion through the entities and attributes that we usually express in text form. Sentiment analysis will classify the polarity of the text in a sentence or document to find out the opinions expressed in the sentence or document are positive or negative.

This research produces k-pop music recommendations from positive and negative comments on trending YouTube using the Naïve Bayes Classifier method. The result of the accuracy of naive Bayes classification is 79% after being tested using the confusion matrix method.

**Keywords:** Sentiment analysis, Naïve bayes classifier, k-pop, Negative comment, Positive comment, YouTube

