

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

PENERAPAN *NAIVE BAYES CLASSIFIER* UNTUK ANALISIS SENTIMEN TERHADAP PENGARUH *KOREAN WAVE* DI INDONESIA PADA YOUTUBE

TUGAS AKHIR

Diajukan Sebagai Salah Satu Syarat
untuk Memperoleh Gelar Sarjana Komputer pada
Program Studi Sistem Informasi

Oleh:

IRINA
11753200071



UIN SUSKA RIAU

FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI SULTAN SYARIF KASIM RIAU
PEKANBARU
2023

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber.
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

LEMBAR PERSETUJUAN

PENERAPAN *NAIVE BAYES CLASSIFIER* UNTUK ANALISIS SENTIMEN TERHADAP PENGARUH *KOREAN WAVE* DI INDONESIA PADA YOUTUBE

TUGAS AKHIR

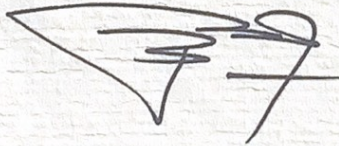
Oleh:

IRINA

11753200071

Telah diperiksa dan disetujui sebagai Laporan Tugas Akhir
di Pekanbaru, pada tanggal 31 Januari 2023

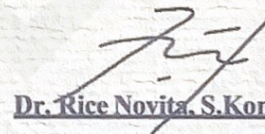
Ketua Program Studi



Eki Saputra, S.Kom., M.Kom.

NIP. 198307162011011008

Pembimbing



Dr. Rice Novita, S.Kom., M.Kom.

NIK. 130510011

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber.

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan satu masalah.

b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

LEMBAR PENGESAHAN

PENERAPAN NAIVE BAYES CLASSIFIER UNTUK ANALISIS SENTIMEN TERHADAP PENGARUH KOREAN WAVE DI INDONESIA PADA YOUTUBE

TUGAS AKHIR

Oleh:

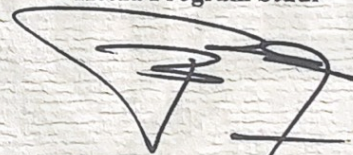
IRINA
11753200071

Telah dipertahankan di depan sidang dewan penguji sebagai salah satu syarat untuk memperoleh gelar Sarjana Komputer Fakultas Sains dan Teknologi Universitas Islam Negeri Sultan Syarif Kasim Riau di Pekanbaru, pada tanggal 27 Januari 2023

Pekanbaru, 27 Januari 2023

Mengesahkan,

Ketua Program Studi



Eki Saputra, S.Kom., M.Kom.
NIP. 198307162011011008



Dekan
Dr. Hartono, M.Pd.
NIP. 196403011992031003

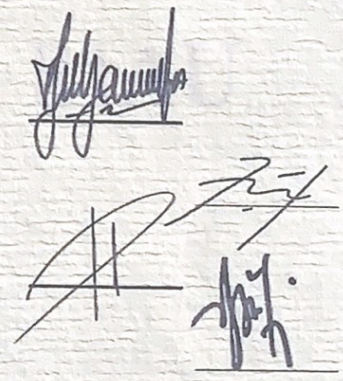
DEWAN PENGUJI:

Ketua : T. Khairil Ahsyar, S.Kom., M.Kom.

Sekretaris : Dr. Rice Novita, S.Kom., M.Kom.

Anggota 1 : Syaifullah, SE., M.Sc.

Anggota 2 : Fitriani Muttakin, S.Kom., M.Cs.



LEMBAR HAK ATAS KEKAYAAN INTELEKTUAL

Tugas Akhir yang tidak diterbitkan ini terdaftar dan tersedia di Perpustakaan Universitas Islam Negeri Sultan Syarif Kasim Riau adalah terbuka untuk umum, dengan ketentuan bahwa hak cipta ada pada penulis. Referensi kepustakaan diperkenankan dicatat, tetapi pengutipan atau ringkasan hanya dapat dilakukan atas izin penulis dan harus dilakukan mengikuti kaedah dan kebiasaan ilmiah serta menyebutkan sumbernya.

Penggandaan atau penerbitan sebagian atau seluruh Tugas Akhir ini harus memperoleh izin tertulis dari Dekan Fakultas Sains dan Teknologi Universitas Islam Negeri Sultan Syarif Kasim Riau. Perpustakaan dapat meminjamkan Tugas Akhir ini untuk anggotanya dengan mengisi nama, tanda peminjaman dan tanggal pinjam pada *form* peminjaman.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Lampiran Surat :
Nomor : Nomor 25/2021
Tanggal : 10 September 2021

SURAT PERNYATAAN

Saya yang bertandatangan di bawah ini:

Nama : IRINA
NIM : 11753200071
Tempat/Tgl. Lahir : SEB-TEMBILAHAN / 11 FEBRUARI 1999
Fakultas/Pascasarjana : SAINS DAN TEKNOLOGI
Prodi : SISTEM INFORMASI
Judul Disertasi/Thesis/Skripsi/Karya Ilmiah lainnya*:
PENERAPAN NAIVE BAYES CLASSIFIER UNTUK ANALISIS
SENTIMEN TERHADAP PENGARUH KOREAN WAVE DI
INDONESIA PADA YOUTUBE

Menyatakan dengan sebenar-benarnya bahwa :

1. Penulisan Disertasi/Thesis/Skripsi/Karya Ilmiah lainnya* dengan judul sebagaimana tersebut di atas adalah hasil pemikiran dan penelitian saya sendiri.
2. Semua kutipan pada karya tulis saya ini sudah disebutkan sumbernya.
3. Oleh karena itu Disertasi/Thesis/Skripsi/Karya Ilmiah lainnya* saya ini, saya nyatakan bebas dari plagiat.
4. Apa bila dikemudian hari terbukti terdapat plagiat dalam penulisan Disertasi/Thesis/Skripsi/(Karya Ilmiah lainnya)* saya tersebut, maka saya bersedia menerima sanksi sesuai peraturan perundang-undangan.

Demikianlah Surat Pernyataan ini saya buat dengan penuh kesadaran dan tanpa paksaan dari pihak manapun juga.

Pekanbaru, 31 JANUARI 2023
Yang membuat pernyataan



IRINA
NIM : 11753200071

Pilih salah satu sesuai jenis karya tulis

LEMBAR PERSEMBAHAN



Alhamdulillah Rabbil Alamiin, puji syukur kepada Allah SWT yang senantiasa memberikan kesehatan fisik dan kesehatan rohani serta nikmat iman dan nikmat islam, sehingga dalam hidup ini kita diajarkan untuk selalu bersyukur termasuk ketika dalam menyelesaikan Tugas Akhir ini.

Tugas Akhir ini saya persembahkan kepada:

1. Bapak Ali, ayahanda tercinta yang sangat saya sayangi, yang telah membesarkan saya, mengorbankan waktu dan tenaganya untuk mencukupi segala kebutuhan saya. Terima kasih yang tak terhingga untuk setiap pengorbanan ayah, semoga ayah selalu sehat dan bahagia amin.
2. Ibu Ikanaria, ibunda tercinta yang telah melahirkan, membesarkan hingga tumbuh dewasa. Tak pernah ada kata mengeluh, hanya kasih sayang, do'a, dan dukungan yang tiada hingga yang diberikan untuk anaknya. Terima kasih yang tiada habisnya untuk ibu tersayang, semoga ibu diberikan kesehatan dan selalu bahagia.
3. Kakak Rahma, adalah kakak yang hebat yang mandiri dan menjadi contoh untuk saya, agar bisa lebih mandiri dan kuat dalam keadaan apapun. Terima kasih kakak yang selalu memberikan saya semangat dan kekuatan.
4. Nenek Itang, nenek yang memberikan saya nasehat dan yang selalu memperhatikan keadaan cucunya, serta mendo'akan yang terbaik. Semoga nenek senantiasa diberikan umur panjang agar saya bisa membahagiakan beliau.
5. Keluarga besar, yang juga memberikan dukungan, do'a serta semangat yang tiada hentinya. Semoga kita dalam lindungan Allah SWT.

KATA PENGANTAR

Assalamu'alaikum Wr.Wb.

Alhamdulillah, puji syukur penulis ucapkan kehadirat Allah SWT yang telah memberikan rahmat dan karunia-Nya sehingga penulis dapat menyelesaikan laporan Tugas Akhir ini dengan topik “Penerapan *Naive Bayes Classifier* untuk Analisis Sentimen Terhadap Pengaruh *Korean Wafe* di Indonesia pada YouTube”. Sholawat serta salam tidak lupa pula penulis panjatkan kepada Nabi Muhammad SAW dengan lafadz *Allahumma Sholli'ala Sayyidina Muhammad Wa'ala Ali Sayyidina Muhammad*. Pada kesempatan ini penulis mengucapkan terima kasih kepada semua pihak yang telah membantu dalam proses pembuatan Tugas Akhir hingga akhirnya terselesaikan laporan Tugas Akhir ini. Terima kasih tersebut penulis ucapkan kepada :

1. Bapak Prof. Dr. Hairunas, M.Ag sebagai Rektor Universitas Islam Negeri Sultan Syarif Kasim Riau.
2. Bapak Dr. Hartono, M.Pd sebagai Dekan Fakultas Sains dan Teknologi, Universitas Islam Negeri Sultan Syarif Kasim Riau.
3. Bapak Eki Saputra, S.Kom., M.Kom sebagai Ketua Program Studi Sistem Informasi Fakultas Sains dan Teknologi, Universitas Islam Negeri Sultan Syarif Kasim Riau.
4. Ibu Siti Monalisa, ST., M.Kom., Sekretaris Program Studi Sistem Informatika sebagai Koordinator Kerja Praktek Program Studi Sistem Informasi Fakultas Sains dan Teknologi, Universitas Islam Negeri Sultan Syarif Kasim Riau.
5. Bapak M. Afdal, ST., M.Kom., sebagai Dosen Pembimbing Akademik yang telah banyak membantu, meluangkan waktu, pikiran dan tenaga serta memberikan semangat dalam proses akademik sampai menyelesaikan Tugas Akhir ini.
6. Ibu Dr. Rice Novita, S.Kom., M.Kom., sebagai Dosen Pembimbing Tugas Akhir yang telah membimbing, meluangkan waktu, pikiran dan tenaga serta memberikan semangat dalam proses penyelesaian Tugas Akhir ini.
7. Bapak Syaifullah, SE., M.Sc., sebagai Dosen Penguji Satu yang telah banyak memberikan arahan, saran serta kontribusi dalam penyelesaian Tugas Akhir ini.
8. Ibu Fitriani Muttakin, S.Kom., M.Cs sebagai Dosen Penguji Dua yang telah banyak memberikan arahan, saran, dan kontribusi dalam penyelesaian Tugas Akhir ini.

Hak Cipta Diindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber.

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

9. Bapak Tengku Khairil Ahsyar, S.Kom., M.Kom., sebagai Koordinator Tugas Akhir yang telah banyak memberikan arahan, saran, dan kontribusi Tugas Akhir ini.
10. Segenap dosen dan karyawan Program Studi Sistem Informasi Fakultas Sains dan Teknologi Universitas Islam Negeri Sultan Syarif Kasim Riau, terima kasih atas ilmu yang diberikan.
11. Keluarga tercinta khususnya Ayah dan Ibu yang telah mencurahkan semua kasih sayang, do'a dan memberikan dukungan secara moril dan materil setiap waktu.
12. Untuk Team '*Puzzle Research Data Teknologi*' yang selalu memberikan ilmu dan motivasi.
13. Keluarga besar SIF D yang telah memberikan banyak petualangan, dukungan, dan semangat untuk menyelesaikan Tugas Akhir.
14. Kepada teman dan sahabat Nurhalimah, Dini Afilla, dan Fauzan Luthfi yang memberikan semangat, motivasi, serta solusi saat mengalami kendala dalam Tugas Akhir.
15. Kepada Doh Kyungsoo yang menjadi salah satu penyemangat di saat mengalami masalah dan kendala dalam Tugas Akhir.
16. Serta semua pihak yang namanya tidak dapat penulis sebutkan satu persatu, yang telah banyak membantu dalam Penyelesaian Tugas Akhir ini.

Semoga segala kebaikan yang telah diberikan kepada penulis bernilai pahala dan amal ibadah disisi Allah SWT, Amin. Penulis menyadari bahwa Tugas Akhir ini masih banyak kekurangan, untuk itu diharapkan kritik dan saran yang bersifat membangun dari semua pihak dapat disampaikan ke email irina7882016@gmail.com. Atas bantuannya saya ucapkan terima kasih.

Wassalamu'alaikum Wr. Wb.

Pekanbaru, 31 Januari 2023

Penulis,

IRINA
NIM. 11753200071



PENERAPAN *NAIVE BAYES CLASSIFIER* UNTUK ANALISIS SENTIMEN TERHADAP PENGARUH *KOREAN WAVE* DI INDONESIA PADA YOUTUBE

IRINA
NIM: 11753200071

Tanggal Sidang: 27 Januari 2023

Periode Wisuda:-

Program Studi Sistem Informasi
Fakultas Sains dan Teknologi
Universitas Islam Negeri Sultan Syarif Kasim Riau
Jl. Soebrantas, No. 155, Pekanbaru

ABSTRAK

Youtube adalah bentuk dari perkembangan teknologi dan menjadi salah satu *platform* video, tempat para konten kreator membagikan berbagai kontennya. Salah satu konten yang kian diminati adalah *Korean Wave*. *Korean Wave* adalah fenomena budaya Korea yang meliputi Drama, Musik, fashion, makanan dan lainnya. Fenomena *Korean Wave* ini meningkat di masa pandemi karena banyak aktivitas dilakukan WFH sehingga orang-orang sering mengakses video *streaming online*. Dampak dari *Korean Wave* di Indonesia seperti munculnya berbagai produk berbau Korea, kuliner dan tempat wisata. Dari dampak *Korean Wave* tersebut menimbulkan opini seperti komentar yang ditinggalkan di salah satu konten Youtube mengenai *Korean Wave*. Dalam hal ini penelitian ini bertujuan untuk mengetahui opini masyarakat khususnya pengguna media Youtube. Opini dibagi menjadi tiga kelas, yaitu negatif, netral dan positif. Dengan menggunakan data teks 7 ribu data komentar dari beberapa video Youtube. Di olah dengan teknik *Text Mining* menggunakan metode *Naïve Bayes Classifier*. Pembagian data menggunakan *K-Fold Cross Validation* dengan 10 split. Didapatkan hasil akurasi algoritma sebesar 62.75%, dengan opini masyarakat lebih cenderung kearah positif, kata yang sering muncul yaitu “suka”, “kpop”, dan “musik”.

Kata Kunci: *Korean Wave, Naïve Bayes Classifier, Youtube.*

- Hak Cipta Dilindungi Undang-Undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
 2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

SENTIMENT ANALYSIS OF THE EFFECT OF THE KOREAN WAVE IN PANDEMIC TIMES USING NAIVE BAYES CLASSIFIER METHOD ON YOUTUBE

IRINA
NIM: 11753200071

Date of Final Exam: January 27th 2023
Graduation Period:-

Department of Information Systems
Faculty of Science and Technology
State Islamic University of Sultan Syarif Kasim Riau
Soebrantas Street, No. 155, Pekanbaru

ABSTRACT

YouTube is a form of technological development and is a video platform where content creators share their content. Korean Wave is one type of content that is becoming increasingly popular. The Korean Wave is a Korean cultural phenomenon that includes drama, music, fashion, food, and more. This Korean Wave phenomenon increased during the pandemic because many activities were carried out by WFH so that people could access online video streaming. The Korean Wave has resulted in the emergence of various Korean-smelling products, as well as culinary and tourist attractions. The Korean Wave's impact sparked debate, as evidenced by comments left on one of YouTube's Korean Wave videos. In this case, this study aims to determine public opinion, especially among YouTube media users. Opinion is divided into three classes: negative, neutral, and positive. We collected 7,000 comment data points from several YouTube videos using text data. Processed with text mining techniques using the Naïve Bayes classifier method. Data sharing using K-fold cross validation with 10 splits With the results of an algorithm accuracy of 62.75% and public opinion tending to be positive, words that often appear are "like," "kpop," and "music."

Keywords: *Korean Wave, Naïve Bayes Classifier, Youtube.*

DAFTAR ISI

LEMBAR PERSETUJUAN	ii
LEMBAR PENGESAHAN	iii
LEMBAR HAK ATAS KEKAYAAN INTELEKTUAL	iv
LEMBAR PERNYATAAN	v
LEMBAR PERSEMBAHAN	vi
KATA PENGANTAR	vii
ABSTRAK	ix
ABSTRACT	x
DAFTAR ISI	xi
DAFTAR GAMBAR	xiii
DAFTAR TABEL	xiv
DAFTAR SINGKATAN	xv
1 PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Batasan Masalah	3
1.4 Tujuan	3
1.5 Manfaat	4
1.6 Sistematika Penulisan	4
2 LANDASAN TEORI	5
2.1 <i>Korean Wave</i>	5
2.2 Youtube	5
2.3 Analisis Sentimen	7
2.4 <i>Data Mining</i>	7
2.5 <i>Text Mining</i>	10
2.6 <i>Text Preprocessing</i>	10

Hak Cipta Diindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

2.7	Algoritma Nazief dan Andriani	11
2.8	Term Frecuency-Inverse Document Frecuency (TF-IDF)	12
2.9	Naive Bayes Classifier (NBC)	13
2.10	Bahasa Pemrograman <i>Python</i>	14
2.11	Penelitian Terdahulu	15
3	METODOLOGI PENELITIAN	17
3.1	Tahap Perencanaan	17
3.2	Tahap Pengumpulan Data	18
3.3	Tahap Analisis dan Hasil	19
3.4	Tahap Dokumentasi	21
4	ANALISA DAN HASIL	22
4.1	Pengumpulan Data	22
4.2	<i>Cleaning</i>	22
4.3	Pelabelan	23
4.4	Tahap <i>Text Preprocesssing</i>	24
4.5	<i>Term Frequency-Invers Document Frequency</i> (TF-IDF)	25
4.6	Pembagian Data Latih dan Data Uji pada NBC	26
4.7	<i>K-Fold Cross Validation</i> NBC	26
4.8	Hasil Prediksi Kelas Data Menggunakan NBC	27
4.9	Analisis Klasifikasi	28
4.10	Visualisasi Data	28
4.11	Hasil Analisis Penelitian	29
5	PENUTUP	31
5.1	Kesimpulan	31
5.2	Saran	31
DAFTAR PUSTAKA		
LAMPIRAN A HASIL PELABELAN		A - 2
LAMPIRAN B DAFTAR STOPWORDS		B - 2
LAMPIRAN C SURAT PERNTATAAN		C - 3

DAFTAR GAMBAR

1.1	Grafik <i>Chart Streaming Online</i> Secara Global	2
2.1	Peringkat Sosial Media yang digunakan di Indonesia	6
2.2	Peringkat Sosial Media yang digunakan di Indonesia	8
2.3	<i>Roadmap</i> Penelitian	16
3.1	Alur Metodologi Penelitian	17
3.2	Alur <i>Text Preprocessing</i>	19
3.3	Alur <i>Naive Bayes Classifier</i>	20
4.1	Hasil Pembagian Data pada NBC dengan <i>K-Fold Cross Validation</i>	26
4.2	Visualisasi Opini Kata Negatif	28
4.3	Visualisasi Opini Kata Netral	29
4.4	Visualisasi Opini Kata Positif	29

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

DAFTAR TABEL

2.1	<i>Channel</i> Youtube	6
4.1	Data Awal	22
4.2	Hasil <i>Cleaning</i>	23
4.3	Hasil Pelabelan	24
4.4	Daftar Kata <i>Stopwords</i>	25
4.5	Hasil <i>Preprocessing Text</i> Data	25
4.6	Hasil Pembobotan TF-IDF	26
4.7	Hasil Pembagian Data Latih dengan <i>K-Fold Cross Validation</i>	27
4.8	Hasil Prediksi Data dengan <i>K-Fold Cross Validation</i>	27
4.9	Hasil Sentimen	28
A.1	Hasil Pelabelan Data Komentar Youtube	A - 2
B.1	<i>Stopword</i>	B - 2

Hak Cipta Dilindungi Undang-Undang

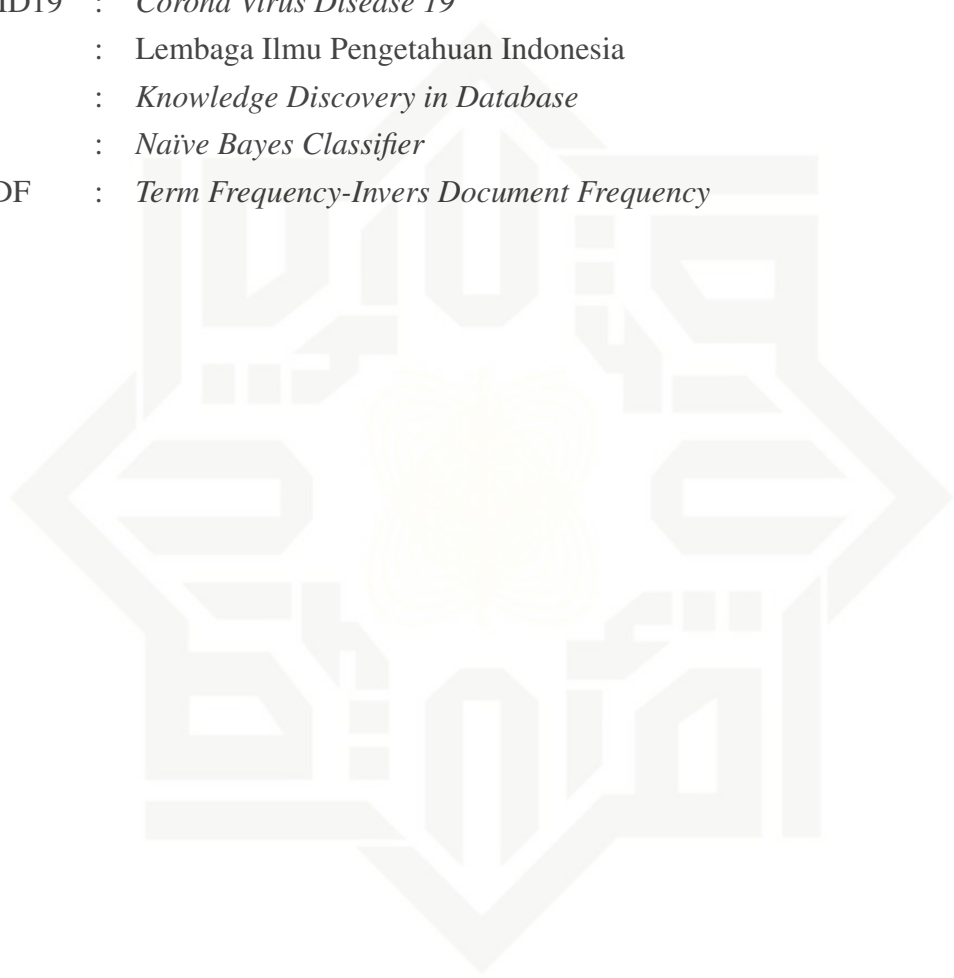
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

DAFTAR SINGKATAN

K-POP	:	<i>Korean Pop</i>
K-Drama	:	<i>Korean Drama</i>
K-Wave	:	<i>Korean Wave</i>
K-Indie	:	<i>Korean Indie</i>
J-Pop	:	<i>Jepang Pop</i>
COVID19	:	<i>Corona Virus Disease 19</i>
LIPI	:	<i>Lembaga Ilmu Pengetahuan Indonesia</i>
KDD	:	<i>Knowledge Discovery in Database</i>
NBC	:	<i>Naïve Bayes Classifier</i>
TF-IDF	:	<i>Term Frequency-Invers Document Frequency</i>



UIN SUSKA RIAU

BAB 1

PENDAHULUAN

1.1 Latar Belakang

Teknologi digital sangat berpengaruh di setiap aktifitas pada zaman globalisasi saat ini, dengan adanya Teknologi digital tidak menjadi hambatan bagi seseorang untuk saling berkomunikasi dan melakukan kegiatan sehari-hari (Mentari, Fauzi, dan Muflikhah, 2018). Perkembangan teknologi digital telah membawa sebuah media baru salah satunya adalah Youtube (Cayari, 2011). Youtube adalah salah satu *platform* video dari Amerika yang telah didirikan oleh Chad Hurley, Steve Chen, dan Jawed Karim pada tahun 2005 (Chandra dkk., 2017). *Platform* media ini menjadi salah satu aplikasi yang paling banyak digunakan oleh pengguna internet di dunia salah satunya negara Indonesia. Berdasarkan data dari *We Are Social* ada 107 juta pengguna YouTube di Indonesia (Datareportal, 2021).

Tentu saja perkembangan penonton Youtube didukung oleh kontribusi para kreator konten atau disebut juga sebagai Youtuber dari Indonesia (Maulida, 2020). Karena fenomena ini membuat banyak orang menemukan cara baru untuk mengekspresikan diri melalui sarana teknologi untuk mengonsumsi, membuat, dan berbagi sebuah karya video (Cayari, 2011). Sehingga semakin banyak tren konten yang diminati oleh penonton Indonesia diantaranya adalah edukasi, memasak, otomotif, *gaming*, bisnis, dan *Korean Pop* dan drama korea (Maulida, 2020). Setiap konten memberikan dampak yang positif maupun negatif sehingga memunculkan komentar-komentar yang bermacam-macam. Dengan demikian, pendapat atau komentar seseorang tentang suatu konten yang dituliskan dalam bahasa yang tertulis tersebut merupakan defenisi sentimen (Oktasari, Chrisnanto, dan Yuniarti, 2016).

Salah satu konten yang kian diminati yakni tentang K-Pop, karena didukung oleh adanya fenomena *Korean Wave* (*K-wave*), dimana ini adalah istilah yang merujuk pada sebuah fenomena budaya Korea mulai dari drama televisi, film, musik, *fashion*, hingga *game*, yang banyak diikuti oleh beberapa negara di Asia (Ahn, Oh, dan Kim, 2013). Tentu saja perkembangan pesat ini didukung dengan adanya sosial media seperti Youtube, Instagram, Twitter dan media sosial lainnya yang membuat K-Pop semakin di gandrungi oleh khalayak luas khususnya untuk anak remaja. Dampak yang ditimbulkan oleh adanya fenomena *Korean Wave* yang semakin berkembang pesat di Indonesia diantaranya munculnya berbagai produk berbau Korea, kuliner dan tempat wisata dengan nuansa Korea, dan masih banyak lagi (Juwita, 2018). Seperti fenomena menarik yang mana para penonton Indonesia memuncaki *chart streaming online* untuk drama Korea secara global (Widi, 2022). Berikut

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

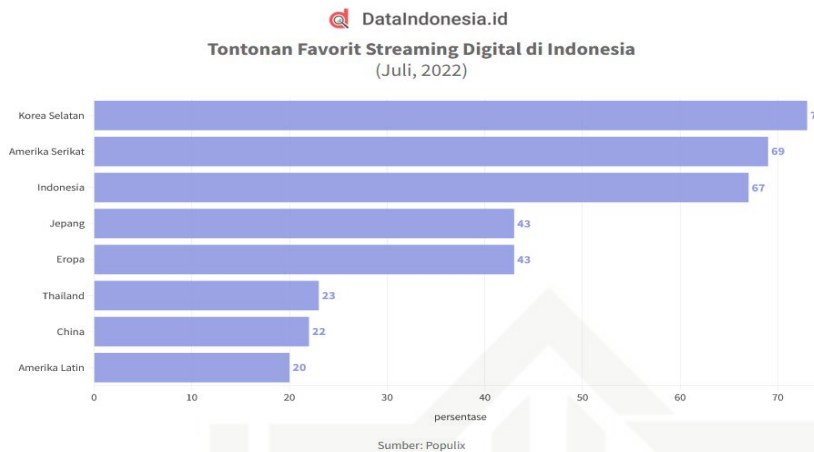
Hak Cipta Diindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

grafik *chart streaming online* secara global dapat dilihat pada Gambar 1.1.



Gambar 1.1. Grafik *Chart Streaming Online* Secara Global

Tampaknya tidak ada yang bisa menghalangi *Korean Wave* di Indonesia. K-Pop, K-Drama, atau bahkan K-Indie musik yang sudah menjadi makanan pokok banyak orang Indonesia. Penggemar budaya pop Korea di Indonesia sangat berkembang pesat. Selain itu, untuk drama dan film Korea, Indonesia juga dapat dengan mudah mengaksesnya disitus web secara gratis dan layanan *streaming video on demand* seperti Youtube. Darynaufal Mulyaman (2020) mengatakan bahwa berdasarkan data Lembaga Ilmu Pengetahuan Indonesia (LIPI), penonton drama Korea di Indonesia telah melonjak dimulai dari pandemi COVID-19. Oleh karena itu, meningkatnya permintaan konten Korea di Indonesia sangat dapat dipahami. Peningkatan ini dibuktikan dengan penggunaan aplikasi *streaming* film, seperti Youtube, Netflix, Viu, dan aplikasi *streaming* lainnya yang didominasi oleh pencarian film dari Korea serta konten-konten musik K-Pop.

Sehingga, tujuan penelitian ini adalah untuk mengetahui bagaimanakah pandangan masyarakat Indonesia terhadap fenomena *Korean Wave*. Dalam penelitian ini hal-hal yang dilihat adalah berapa banyak yang mendukung, menolak atau memilih untuk acuh terhadap adanya fenomena *Korean Wave* ini melalui data yang di ambil dari komentar Youtube dengan teknik *Text Mining* menggunakan metode *Naïve Bayes Classifier* (NBC). NBC merupakan metode sederhana tetapi memiliki akurasi serta performansi yang tinggi dalam pengklasifikasian teks (Ling, Kencana, dan Oka, 2014).

Berdasarkan penelitian sebelumnya, penerapan Text Mining telah dilakukan oleh Syarifuddin (2020) tentang Analisis Sentimen Opini Publik Mengenai COVID-19 pada Twitter Menggunakan Metode *Naïve Bayes Classifier* dan KNN dengan hasil penelitian metode NBC memiliki nilai lebih tinggi, salah satunya adalah t-

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

tingkat akurasi sebesar 63.21% dibandingkan metode KNN yang memiliki akurasi 58.10%. Selanjutnya yaitu penelitian Nurmawati, dkk (2020) tentang Analisis Sentimen Terhadap Penggemar K-Pop di Media Sosial Twitter menggunakan *Naïve Bayes Classifier* (Studi Kasus Penggemar Grup BTS) dengan tingkat akurasi sebesar 75%. Dari analisis yang dilakukan menggunakan klasifikasi algoritma *Naïve Bayes Classifier* terdapat polaritas sentimen negatif sebanyak 34.2%, netral 58.5%, dan positif 7.3%.

Berdasarkan uraian di atas dan didukung oleh penelitian sebelumnya, maka pada penelitian Tugas Akhir ini akan dilakukan penerapan *Naïve Bayes Classifier* untuk analisis sentimen terhadap pengaruh *Korean Wave* di Indonesia pada Youtube.

1.2 Rumusan Masalah

Berdasarkan latar belakang masalah yang telah diuraikan diatas, maka rumusan masalah pada penelitian ini adalah “Bagaimana penerapan metode *Naïve Bayes Classifier* dalam menganalisa sentimen masyarakat terhadap pengaruh *Korean Wave* di Indonesia pada Youtube”.

1.3 Batasan Masalah

Dalam melakukan suatu penelitian diperlukan batasan-batasan agar tidak menyimpang dari apa yang telah direncanakan, adapun batasan masalah dalam penulisan Tugas Akhir ini, yaitu:

1. Data komentar secara garis besar hanya membahas K-Pop dan Drama Korea.
2. Data sentimen dikumpulkan dari komentar di beberapa *channel* Youtube yaitu, Korea Reomit, Narasi Newsroom, dan Fuadh Naim. Indikator *channel* ini dilihat dari jumlah *subscribe*, *like*, dan *views* dari video.
3. Tidak menganalisis kata-kata yang tidak diperlukan seperti *hashtag*, *username* dan *url*.
4. Algoritma yang digunakan untuk pengklasifikasian adalah *Naïve Bayes Classifier*.
5. *Tools* yang digunakan untuk klasifikasi data adalah bahasa pemrograman *Phyton* dan *Microsoft Excel*.
6. Klasifikasi opini menjadi 3 kategori kelas yaitu positif, negatif, dan netral.

1.4 Tujuan

Adapun tujuan yang ingin dicapai oleh penulis dari penelitian ini adalah:

1. Menguji tingkat akurasi metode *Naïve Bayes Classifier* pada Analisis Sentimen *Korean Wave*.

Hak Cipta Diindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

2. Mengetahui opini masyarakat terhadap fenomena *Korean Wave* di Indonesia.
3. Mengvisualisasikan komentar dari data Youtube berupa kata-kata yang sering muncul berdasarkan masing-masing sentimen.

1.5 Manfaat

Manfaat yang diperoleh dari penelitian ini adalah:

1. Memberikan informasi mengenai tingkat akurasi yang digunakan pada NBC untuk dikembangkan pada kasus *Text Mining*.
2. Memberikan informasi mengenai opini masyarakat terhadap fenomena *Korean Wave* di Indonesia.
3. Mendapatkan visualisasi dari komentar berupa kata-kata yang sering muncul berdasarkan masing-masing kelas sentimen guna analisis lebih lanjut.

1.6 Sistematika Penulisan

Sistematika penulisan laporan Tugas Akhir ini disusun untuk memberikan gambaran umum tentang penelitian yang penulis lakukan. Berikut merupakan sistematika penulisan pada laporan Tugas Akhir yang terdiri dari 5 (lima) bab, yaitu:

BAB 1. PENDAHULUAN

Bab ini berisi gambaran umum dari Tugas Akhir ini, yang meliputi latar belakang permasalahan, rumusan masalah, batasan masalah, tujuan dari pembahasan dan manfaat yang didapatkan.

BAB 2. LANDASAN TEORI

Bab ini berisi teori-teori yang berasal dari jurnal, buku, serta studi kepustakaan yang digunakan sebagai landasan teori dalam pembuatan laporan Tugas Akhir ini.

BAB 3. METODOLOGI PENELITIAN

Bab ini berisi tentang metodologi atau alur penelitian yang digunakan dalam penyusunan Tugas Akhir ini.

BAB 4. ANALISA DAN HASIL

Bab ini membahas tentang analisis komentar pada video Youtube tentang *Korean Wave* di beberapa *channel* Youtube dan mengklasifikasi komentar menggunakan algoritma *Naïve Bayes Classifier*.

BAB 5. PENUTUP

Bab ini berisikan kesimpulan dari Tugas Akhir yang dibuat dan saran untuk penelitian selanjutnya.



BAB 2

LANDASAN TEORI

2.1 Korean Wave

Korean Wave (K-Wave) merupakan istilah yang merujuk pada budaya Korea yang tersebar secara global melalui drama televisi, film, lagu, dan selebriti terkait yang telah mendapatkan popularitas yang luar biasa di China, Taiwan, Hong Kong, Negara-negara Asia Timur dan Tenggara lainnya (Shim, 2006). Tak terkecuali di Indonesia, di mana fenomena K-Wave ini digencarkan melalui drama Korea hingga *music pop* Korea (K-Pop).

Musik Korea ini mampu memenuhi selera musik dunia khususnya Amerika Serikat, sehingga musik Korea mampu mendominasi pasar musik dunia. Meski ada juga J-Pop (musik pop Jepang) dan aliran musik lainnya. Musik pop Korea lebih mampu memuaskan selera generasi sekarang, sehingga K-Pop lebih digemari (Rahmawati, Nikmah, Perwira, dan Rakhmawati, 2021). Sering para remaja begitu terobsesi dengan hal-hal yang berbau Korea. Sehingga mereka tidak mengetahui perkembangan dunia *entertainment* Indonesia, justru mereka lebih mengikuti dan antusias dengan perkemabangan dunia *entertainment* Korea, banyak dari mereka memilih *fashion* ala Korea, mempelajari tulisan dan bahasa Korea tetapi tidak pernah mempelajari bahasa daerahnya sendiri. Bahkan di jejaring sosial banyak dari mereka yang mencantumkan beberapa kota di Korea Selatan sebagai kota tempat tinggal atau kota asalnya, banyak juga yang menggunakan nama Korea untuk nama jejaring sosial mereka. Tentunya hal ini memberikan dampak bagi pembentukan mental remaja Indonesia.

2.2 Youtube

Youtube merupakan media sosial yang telah mengalami perkembangan begitu luar biasa pada saat ini, Youtube merupakan media sosial untuk berbagi video yang terbentuk pada tahun 2005 sebelum pada akhirnya dimiliki oleh Google pada tahun 2006 sampai sekarang yang didirikan oleh Chad Hurley, Steve Chen, Jawed Karim mereka merupakan tiga mantan karyawan perusahaan *paypal*.

Youtube tidak hanya berkembang sebagai suatu media tempat berbagi video dari para penggunanya, namun juga berkembang menjadi suatu bentuk media sosial yang dinamis dan menjadi salah satu media yang paling populer di Inggris (Misoch, 2014). Indonesia termasuk salah satu negara dengan pengguna aktif Youtube terbesar di dunia. Dalam rilis *Datareportal* disebutkan Youtube menduduki peringkat pertama sosial media yang sering digunakan di Indonesia pada April 2021. Berikut

Hak Cipta Dilindungi Undang-Undang

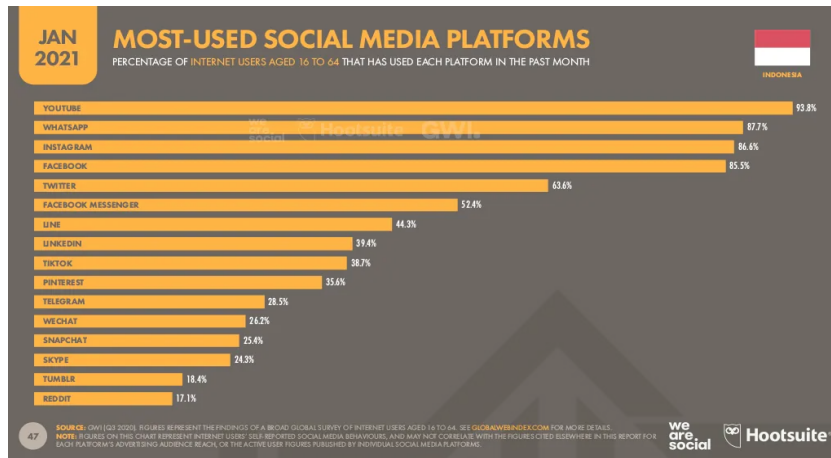
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber.

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

peringkat sosial media yang digunakan di Indonesia dapat dilihat pada Gambar 2.1.



Gambar 2.1. Peringkat Sosial Media yang digunakan di Indonesia

Konten yang ada pada Youtube memiliki beragam pilihan, seperti *music*, *trailer film*, *fashion*, *vlog*, *games*, dan masih banyak yang lainnya. Para pengguna Youtube dapat melihat informasi jumlah *view* dan jumlah berapa yang menyukai dan tidak menyukai video tersebut, Youtube juga memiliki fitur untuk berbagi komentar sebagai akses komunikasi untuk pengguna kepada pemilik akun Youtube. Ada tiga *Channel* Youtube yang digunakan pada penelitian ini dan beberapa video dari setiap *channel* yang akan digunakan sebagai data pada penelitian ini. Berikut *channel* Youtube yang digunakan beserta link video pada Tabel 2.1.

Tabel 2.1. Channel Youtube

No.	Channel	Subscriber	Link video
1.	Korea Reomit	4,84 juta	https://youtu.be/M2joY614Ce8
2.	Narasi Newsroom	557 ribu	https://youtu.be/iKGFyyUN_eg https://youtu.be/q0Uwy_fdMP0
3.	Fuadh Naim	72,1 ribu	https://youtu.be/6Pv_G3LMD4A https://youtu.be/lxhZER8abiw

1. Korea Reomit adalah *channel* dari seorang Youtuber bernama Jang Han-sol berkebangsaan Korea Selatan yang pindah ke Indonesia mulai dari usia empat tahun. Ia membuat kanal Youtube tersebut untuk membangun citra yang baik untuk Indonesia dan Korea melalui pembahasan seputar kebudayaan dua bangsa tersebut.
2. Narasi Newsroom adalah *channel* Youtube bagian dari Narasi yang di dirikan oleh Najwa Shihab pada awal tahun 2018. Narasi juga memproduksi konten *jurnalisme* dalam bentuk video, perbincangan, *reportase*, kemasn

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber.
2. Dilarang mengutip sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

documenter, opini dan ruang interaksi, dan sebagainya.

3. Fuadh Naim adalah *channel* dari seorang Youtuber yang bernama Fuadh Naim, ia adalah seorang *trainer* di YukNgaji, Director di Underblack Picture dan Penulis Buku yang berjudul “Pernah Tenggelam” sekaligus pendakwah melalui media sosial. Sebelumnya ia adalah seorang K-Wavers yang mencintai segala yang bersangkutan dengan budaya Korea.

2.3 Analisis Sentimen

Analisis sentimen adalah sebuah proses untuk menentukan sentimen atau opini dari seseorang yang diwujudkan dalam bentuk teks dan bisa dikategorikan sebagai sentimen positif atau negatif (Hadna, Santosa, dan Winarno, 2016). *Sentiment analysis* atau *opinion mining* mengacu pada bidang yang luas dari pengolahan bahasa alami, komputasi linguistik dan *Text Mining* yang bertujuan menganalisa pendapat, sentimen, evaluasi, sikap, penilaian dan emosi seseorang apakah pembicara atau penulis berkenaan dengan suatu topik, produk, layanan, organisasi, individu, maupun kegiatan tertentu (Lin, Wang, dan Zhou, 2016).

Pengguna internet banyak menuliskan pengalaman, opini dan segala hal yang menjadi perhatian mereka. Tulisan tentang apa yang mereka rasakan ini berupa perasaan positif, netral, maupun negatif yang bisa diungkapkan dengan cara yang cukup kompleks. Menurut Manalu (2014) hal ini memungkinkan kita untuk mencari informasi tentang:

1. Deteksi *Flame*
2. Persepsi produk baru
3. Persepsi Merek
4. Manajemen reputasi

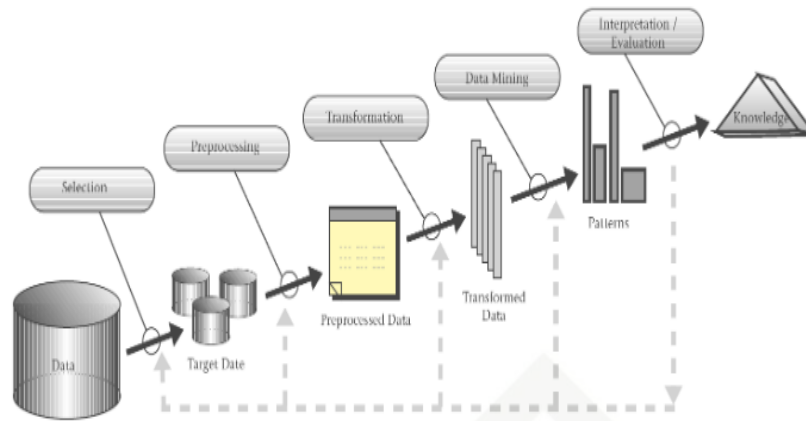
2.4 Data Mining

Data Mining merupakan proses menemukan pola dan pengetahuan menarik dari sejumlah besar data. Sumber data mencakup database, gudang data, web, repositori informasi lain, atau data yang dialirkan ke sistem secara dinamis. *Data Mining* digunakan untuk mengekstrak informasi yang didapatkan dalam *database* dari data mentah (Ahmed dan Jesmin, 2014).

Dalam *Data Mining*, muncul istilah ‘*Knowledge Discovery in Database*’ atau KDD yang meliputi pengumpulan, klasifikasi dan evaluasi data yang relevan (Mukherjee, Shaw, Haldar, dan Changdar, 2015). Proses dalam KDD bersifat interaktif dan berulang, yang melibatkan banyak langkah dengan banyak keputusan yang dibuat oleh pengguna (Fayyad, Piatetsky-Shapiro, dan Smyth, 1996). Adapun tahapan menurut Fayyad dkk. (1996) dalam KDD dapat dilihat pada Gambar 2.2

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.



Gambar 2.2. Peringkat Sosial Media yang digunakan di Indonesia

1. *Data Cleaning*

Data-data yang tidak relevan lebih baik dibersihkan karena keberadaannya bisa mengurangi mutu atau akurasi dari hasil *Data Mining* merupakan istilah yang sering dipakai untuk menggambarkan tahapan ini. pembersihan data juga dapat mempengaruhi performansi dari sistem *Data Mining* karena data yang ditangani akan berkurang jumlah dan kompleksitasnya.

2. *Data Integration*

Data Mining tidak hanya berasal dari suatu *database* tetapi juga berasal dari beberapa *database* atau file teks.

3. *Data Transformation*

Coding adalah proses transformasi pada data yang telah dipilih, sehingga data tersebut sesuai untuk proses *Data Mining*. Proses *coding* dalam KDD merupakan proses kreatif dan sangat tergantung pada jenis atau pola informasi yang akan dicari dalam basis data.

4. *Data Mining*

Data Mining adalah proses pencarian pola atau informasi menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu. Teknik, metode, atau algoritma dalam *Data Mining* sangat bervariasi. Pemilihan metode atau algoritma yang tepat sangat bergantung pada tujuan dan proses KDD secara keseluruhan.

5. *Pattern Evolution*

Dalam tahap ini hasil dari teknik *Data Mining* berupa pola-pola yang khas maupun model prediksi dievaluasi untuk menilai apakah hipotesa yang ada memang tercapai. Bila ternyata hasil yang diperoleh tidak sesuai hipotesa, ada beberapa alternatif yang bisa diambil seperti menjadikannya umpan

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

balik untuk memperbaiki proses *Data Mining*, menggunakan teknik *Data Mining* yang lain yang lebih sesuai. atau menerima hasil ini sebagai suatu hasil yang diluar dugaan yang mungkin bermanfaat. Ada beberapa teknik *Data Mining* yang menghasilkan hasil analisa berjumlah besar seperti analisa prediksi. Visualisasi hasil analisis akan sangat membantu untuk memudahkan pemahaman dari hasil *Data Mining*.

6. *Knowledge*

Tahap terakhir dari proses *Data Mining* adalah bagaimana memformulasikan keputusan atau aksi dari hasil analisa yang didapat. Ada kalanya hal ini harus melibatkan orang-orang yang tidak memahami *Data Mining*. Karenanya presentasi hasil *Data Mining* dalam bentuk pengetahuan yang bisa dipahami semua orang adalah satu tahapan yang diperlukan dalam proses *Data Mining*.

Menurut Fayyad, dkk. (1996) teknik yang digunakan dalam *Data Mining* adalah sebagai berikut:

1. *Association Rules*

Merupakan teknik dalam *Data Mining* yang sudah familiar dan mudah untuk dipahami. Mekanisme yang berjalan pada algoritma ini adalah membandingkan antara dua item atau lebih yang kemudian dapat dianalisa pola kemiripan antar keduanya, misalnya pola pembelian pelanggan terhadap barang-barang tertentu.

2. *Classification*

Klasifikasi dapat digunakan untuk berbagai hal, mengidentifikasi kelas tertentu misalnya mengklasikasikan mobil ke dalam beberapa jenis, teknik ini juga dapat dikombinasikan dengan teknik yang lain seperti *decision tree* dan *clustering*.

3. *Clustering*

Teknik ini bertujuan untuk mengidentifikasi informasi yang berkorelasi namun pada umumnya berbeda sehingga diantaranya terdapat garis yang bersinggungan, atau dengan kata lain hubungan yang tidak terduga, sehingga menghasilkan informasi yang bermanfaat.

4. *Prediction*

Salah satu teknik inti ini hampir mirip dengan teknik *Data Mining classification* letak perbedaannya hanya pada penekanannya. Fungsi prediksi ini berbeda karena catatan diklasifikasikan menurut beberapa prakiraan perilaku.

2.5 Text Mining

Text Mining merupakan salah satu bidang khusus dari *Data Mining*. *Data Mining* adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan *machine learning* untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai *database* besar (Turban, 2005). *Text Mining* adalah proses yang digunakan untuk mengekstrak pola yang menarik dan signifikan untuk mengeksplorasi pengetahuan dari sumber data tekstual. Beberapa teknik *Text Mining* seperti peringkasan, klasifikasi, *clustering* dan lainnya dapat diterapkan untuk mengekstrak pengetahuan (Talib, Hanif, Ayesha, dan Fatima, 2016).

2.6 Text Preprocessing

Text Preprocessing merupakan tahap yang awal dalam *Text Mining* (Efendi dan Mustakim, 2017). Tahap *Text Preprocessing* dilakukan untuk mengubah data tekstual yang tidak terstruktur menjadi data yang terstruktur. Ada beberapa langkah yang dilakukan di *Text Preprocessing*. Secara umum tahapan dalam *Text Preprocessing* dibagi menjadi dua bagian utama, yang pertama adalah *Tokenizing*, *Case Folding*, *Spelling Normalization*, dan *filtering*. Sedangkan yang kedua adalah *Stemming*. Berikut penjelasan dari *Text Preprocessing* menurut Surjandari, dkk. (2016), Yanis dan Iriani, (2018):

1. *Tokenizing*

Tokenizing adalah proses memisahkan dokumen menjadi potongan-potongan kata tunggal.

2. *Case Folding*

Case Folding adalah proses megubah semua huruf besar menjadi huruf kecil dalam dokumen dan sebaliknya. Langkah ini juga untuk memastikan bahwa hanya huruf a sampai z yang terdapat dalam dokumen.

3. *Spelling Normalization*

Spelling Normalization adalah proses mengganti kata yang salah dalam ejaan atau disingkat. Misalnya kata “tidak” memiliki beberapa bentuk lain, seperti “tdk”, “gak”, “nggak”, “enggak” (yang artinya bukan dalam bahasa Indonesia).

4. *Filtering*

Filtering adalah proses menghilangkan kata-kata yang sering muncul tetapi tidak mempunyai efek apapun dalam ekstraksi klasifikasi teks. Dalam proses ini, kata-kata yang disertakan seperti timer, kata tanya, dan konjungsi.

5. *Stemming*

Hak Cipta Diindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Stemming adalah proses pemetaan dan penguraian bentuk kata menjadi bentuk dasarnya. Tujuan dari proses stemming adalah untuk menghilangkan afiks, baik dalam bentuk *prefix*, *suffix*, atau *confix* yang ada disetiap kata.

2.7 Algoritma Nazief dan Andriani

Algoritma Nazief dan Andriani dikembangkan pertama kali oleh Bobby Nazief dan Mirna Andriani. Algoritma ini berdasarkan pada aturan morfologi bahasa Indonesia yang luas, yang dikumpulkan menjadi satu grup dan di-enkapsulasi pada imbuhan/*affixes* yang diperbolehkan (*allowed affixes*) dan imbuhan/*affixes* yang tidak diperbolehkan (*disallowed affixes*). Algoritma ini menggunakan kamus kata dasar dan mendukung *recoding*, yakni penyusunan kembali kata-kata yang mengalami proses *stemming* berlebih.

Menurut D. Nugroho, (2017) langkah-langkah algoritma Nazief dan Andriani adalah:

1. Kata yang belum di *stemming* dicari pada kamus, jika ditemukan, kata tersebut dianggap sebagai kata dasar yang benar dan algoritma dihentikan.
2. Hilangkan *Inflectional suffixes*, yaitu dengan menghilangkan *particle* (“-lah”, ”-kah”, “-tah” atau “-pun”), kemudian hilangkan *inflectional possessive pronoun suffixes* (“-ku”, “-mu” atau ”-nya”). Cek kata di dalam kamus kata dasar, jika ditemukan, algoritma dihentikan, jika tidak lanjut ke langkah 3.
3. Hapus *Derivational Suffix* (“-i” atau ”-an”,”). Jika kata ditemukan dalam kamus kata dasar, maka algoritma berhenti. Jika tidak, maka lanjut ke langkah berikut:
 - (a) Jika akhiran “-an” telah dihapus dan huruf terakhir dari kata tersebut adalah “-k”, maka “-k” juga dihapus. Jika kata tersebut ditemukan dalam kamus maka algoritma berhenti. Jika tidak ditemukan maka lakukan langkah selanjutnya.
 - (b) Akhiran yang dihapus (“-i”, “- an” atau “-kan”) dikembalikan, lanjut ke langkah 4.
4. Hapus *Derivational Prefix* (“be-”, ”di-”, ”ke-”, ”me-”, ”pe-”, ”se-” dan “te-“). Jika kata yang didapat ditemukan didalam database kata dasar, maka proses dihentikan, jika tidak, maka lakukan *recoding*. Tahapan ini dihentikan jika memenuhi beberapa kondisi berikut:
 - (a) Terdapat kombinasi awalan dan akhiran yang tidak di izinkan.
 - (b) Awalan yang dideteksi sama dengan awalan yang dihilangkan sebelumnya.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

(c) Tiga awalan telah dihilangkan.

5. Jika semua langkah telah dilakukan tetapi kata dasar tersebut tidak ditemukan pada kamus, maka algoritma ini mengembalikan kata yang asli sebelum dilakukan *stemming*.

Dalam penelitiannya Asian, dkk. (2017) melakukan beberapa pengembangan algoritma Nazief dan Adriani sebagai berikut:

1. Menggunakan kamus kata yang lebih lengkap.
2. Menambahkan aturan-aturan untuk kata-kata majemuk perulangan.
3. Menambahkan aturan awalan dan akhiran, serta aturan lainnya, yaitu:
 - (a) Menambahkan partikel (*inflection suffix*) “-pun”.
 - (b) Penambahan aturan pemenggalan awalan.
 - (c) Perubahan aturan pemenggalan untuk tipe awalan “me”.
4. Perubahan urutan proses stemming, yaitu:
 - (a) Kata dengan awalan “be-” dan akhiran “-lah”, hilangkan awalan terlebih dahulu kemudian akhiran.
 - (b) Kata dengan awalan “be-” dan akhiran “-an”, hilangkan awalan terlebih dahulu kemudian akhiran.
 - (c) Kata dengan awalan “me-“ dan akhiran “-i”, hilangkan awalan terlebih dahulu kemudian akhiran.
 - (d) Kata dengan awalan “di-“ dan akhiran “-i”, hilangkan awalan terlebih dahulu kemudian akhiran.
 - (e) Kata dengan awalan “pe-“ dan akhiran “-i”, hilangkan awalan terlebih dahulu kemudian akhiran.
 - (f) Kata dengan awalan “ter-“ dan akhiran “-i”, hilangkan awalan terlebih dahulu kemudian akhiran (Agusta, Kristen, dan Wacana, 2009).

2.8 Term Frequency-Inverse Document Frequency (TF-IDF)

TF-IDF sering digunakan sebagai faktor pembobotan dalam information retrieval dan text mining. Nilai TF-IDF meningkat secara proporsional berdasarkan berapa banyak kemunculan kata dalam dokumen (*term frequency*), tetapi dinetralkan oleh frekuensi kata dalam corpus (*inverse document frequency*) (Vijayarani, Ilamathi, Nithya, dkk., 2015). Menurut Nurjannah, dkk. (2017) tahapan pada pembobotan kata yaitu sebagai berikut:

1. *Term Frequency* (TF)

TF merupakan frekuensi kemunculan kata pada suatu dokumen teks. TF didefinisikan jumlah kemunculan term t pada dokumen d . Persamaan dari *Term Frequency* tf_{td} ditunjukkan pada Persamaan 2.1.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

$$tf = 0.5 + 0.5 \frac{tf}{\max(tf)} \quad (2.1)$$

Keterangan:

tf adalah banyaknya kata yang dicari pada sebuah dokumen.

$\max(tf)$ adalah jumlah kemunculan terbanyak term pada dokumen yang sama.

2. *Document Frequency (DF)*

DF merupakan kata-kata yang banyak terdapat pada dokumen, kata tersebut tidak informatif, seperti kata dan, di, atau, bisa, merupakan.

3. *Invers Document Frequency (IDF)*

IDF merupakan frekuensi kemunculan term pada keseluruhan dokumen teks. *Term* yang jarang muncul pada keseluruhan dokumen teks memiliki nilai IDF lebih besar dibandingkan dengan term yang sering muncul. Persamaan dari IDF ditunjukkan pada Persamaan 2.2.

$$idf_t = \log\left(\frac{N}{df_t}\right) \quad (2.2)$$

Keterangan:

N adalah jumlah dokumen teks.

df_t adalah jumlah dokumen yang mengandung *term t*.

4. *Term Frequency – Invers Document Frequency (TF-IDF)*

Nilai TF-IDF dari sebuah kata merupakan kombinasi dari nilai TF dan nilai IDF dalam perhitungan bobot. Persamaan dari TF-IDF ditunjukkan pada Persamaan 2.3.

$$W_{td} = W_{tfd} \times idf_t \quad (2.3)$$

Keterangan:

W_{tfd} adalah *Term Frequency*.

idf_t adalah *Invers Document Frequency*.

2.9 Naive Bayes Classifier (NBC)

Algoritma NBC merupakan algoritma yang digunakan untuk mencari nilai probabilitas tertinggi untuk mengklasifikasi data uji pada kategori yang paling tepat (Feldman dan Sanger, 2007). NBC adalah salah satu algoritma yang populer digunakan untuk keperluan *Data Mining* karena kemudahan penggunaannya serta waktu pemrosesan yang cepat, mudah diimplementasikan dengan 22 strukturnya yang cukup sederhana dan tingkat efektifitas yang tinggi (Taheri dan Mammadov,

2013).

NBC adalah metode klasifikasi yang berdasarkan probabilitas dan Teorema Bayesian dengan asumsi bahwa setiap variable X bersifat bebas (*independence*). Dengan kata lain, NBC mengasumsikan bahwa keberadaan sebuah atribut (*variable*) tidak ada kaitannya dengan keberadaan atribut (*variable*) yang lain. Berikut merupakan rumus perhitungan NBC, dapat dilihat pada Persamaan 2.4, Persamaan 2.5, dan Persamaan 2.6 (Jurafsky dan Martin, 2014).

$$P(c) = \frac{N_c}{N} \quad (2.4)$$

Keterangan:

c adalah Kelas.

N_c adalah Banyak dokumen dalam suatu kelas n . N adalah Jumlah keseluruhan data latih.

$$P(W|j_c) = \frac{\text{count}(w, c) + 1}{\text{count}(c) + j_j} \quad (2.5)$$

Keterangan:

$\text{count}(w, c)$ adalah Frekuensi kata w pada kelas c .

$\text{count}(c)$ adalah Total frekuensi pada masing-masing kelas c .

j_j adalah Total kata unik pada keseluruhan kelas c .

$$P_{(c|jdn)} = P(c) \mathbf{O}_{P(w|j_c)} \quad (2.6)$$

Keterangan:

$P_{c|jdn}$ adalah *Choosing a class*.

$P(c)$ adalah *Priors*.

$P_{P(w|j_c)}$ adalah *Total conditional probabilities*.

2.10 Bahasa Pemrograman Python

Menurut Nosrati (2011) *Python* adalah bahasa pemrograman yang fleksibel dan sederhana yang didefinisikan dalam dokumen-dokumennya sebagai berikut:

1. *Python* adalah bahasa pemrograman tujuan umum yang sangat tingkat tinggi, dinamis, berorientasi objek, yang umum digunakan.
2. Bisa digunakan dalam aplikasi yang luas.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

3. Bahasa ini dapat mendukung berbagai gaya pemrograman termasuk struktural dan berorientasi objek. Gaya lain juga bisa digunakan.
4. *Python* sangat fleksibel, karena kemampuannya untuk menggunakan komponen modular yang dirancang dalam bahasa pemrograman lainnya. Sebagai contoh, Anda dapat menulis sebuah program di C++ dan mengimpornya ke *python* sebagai modul.

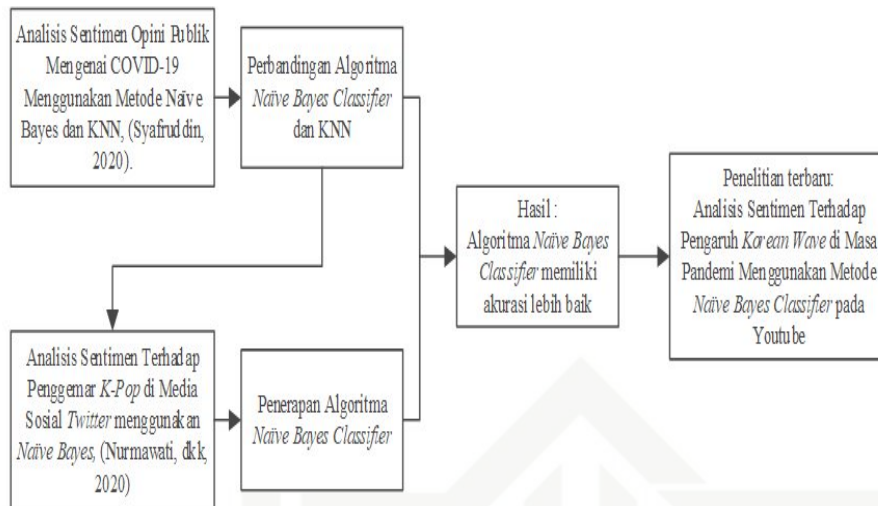
2.11 Penelitian Terdahulu

Beberapa penelitian yang terkait dengan penelitian Tugas Akhir ini diantaranya adalah penelitian tentang Analisis Sentimen Opini Publik Mengena i COVID-19 Menggunakan Metode *Naïve Bayes Classifier* dan KNN oleh Syarifuddin (2020). Proses penelitian ini menggunakan API Twitter sebanyak 1098 opini dengan kata kunci “COVID-19”, dimulai dari klasifikasi opini positif atau negatif, data *cleaning, preprocessing*, hingga di dapatkan hasil akhir. Dari hasil pengujian diketahui metode *Naïve Bayes Classifier* memiliki nilai lebih tinggi, salah satunya adalah tingkat accuracy sebesar 63.21%, sedangkan metode KNN sebesar 58.10%, dan didapatkan pula kecenderungan opini masyarakat di Twitter condong positif, hal tersebut dapat dilihat dari jumlah opini positif sebesar 610 sedangkan negatif 488, ditunjang dengan hasil pengujian *precision* di metode *Naïve Bayes Classifier* dengan nilai positif lebih tinggi dari pada negatif yaitu 66.40%: 58.94%.

Selanjutnya yaitu penelitian tentang Analisis Sentimen Terhadap Penggemar K-Pop di Media Sosial Twitter menggunakan *Naïve Bayes Classifier* (Studi Kasus Penggemar Grup BTS). Dari analisis yang dilakukan menggunakan klasifikasi algoritma *Naïve Bayes Classifier* terdapat polaritas sentimen negatif sebanyak 34.2%, netral 58.5%, dan positif 7,3%. dari 1000 data yang diambil sesuai hasil popularitas *tweet* yang bernilai negatif sesuai hasil popularitas sebanyak 342. Dengan tingkat akurasi sebesar 75%. Roadmap penelitian sebelumnya terhadap penelitian yang peneliti lakukan dapat dilihat pada Gambar 2.3.

Hak Cipta Diindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.



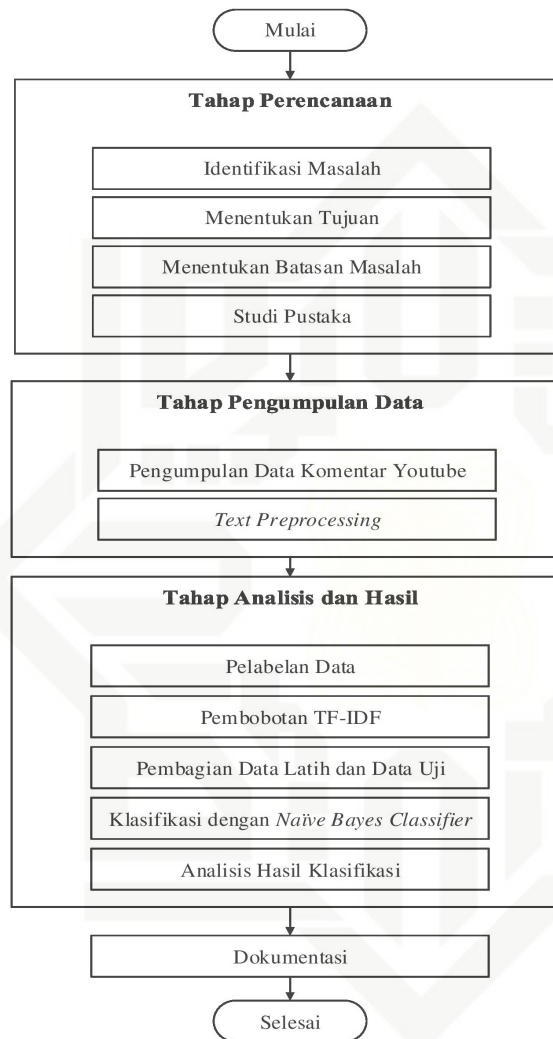
Gambar 2.3. Roadmap Penelitian

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

BAB 3 METODOLOGI PENELITIAN

Berikut alur penelitian yang akan dilakukan dalam Tugas Akhir ini dapat dilihat pada Gambar 3.1.



Gambar 3.1. Alur Metodologi Penelitian

Berikut ini merupakan penjelasan dari langkah-langkah metodologi penelitian Tugas Akhir.

3.1 Tahap Perencanaan

Pada kegiatan ini peneliti menjelaskan permasalahan sebagai latar belakang penelitian dan tujuan yang akan dicapai. Permasalahan dalam penelitian ini berasal dari studi literatur yang telah dipelajari oleh peneliti dan akan diimplementasikan pada data komentar pada video Youtube.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

1. Identifikasi Masalah
Sebelum memulai penelitian maka penulis mengidentifikasi permasalahan yang terjadi pada topik penelitiann yang akan dilakukan.
2. Menentukan Tujuan Penelitian
Untuk mendukung tercapainya sasaran Tugas Akhir ini, tahapan selanjutnya adalah penentuan tujuan dari Tugas Akhir, agar tujuan dalam penulisan ini lebih terfokuskan. Tujuan dari penelitian ini adalah untuk mengetahui pandangan masyarakat terhadap fenomena *Korean Wave* berdasarkan sentimen masyarakat melalui komentar video Youtube menggunakan metode NBC.
3. Menentukan Batasan Masalah
Penentuan batasan masalah bertujuan agar penelitian tidak melebar atau keluar dari topik penelitian.
4. Studi Pustaka
Studi pustaka yaitu melakukan pengumpulan data sebagai landasan teori yang kuat untuk melakukan penyusunan laporan dan menganalisa permasalahan. Studi pustaka yang dilakukan pada penelitian ini yaitu dengan memanfaatkan literatur dari perpustakaan sebagai dasar dari penelitian lapangan, buku-buku penunjang yang berupa *text book* maupun *e-book*, jurnal internasional dan nasional yang berhubungan dengan topik yang diteliti, skripsi, serta beberapa *website* resmi untuk mendukung penelitian ini.

3.2 Tahap Pengumpulan Data

Tahap ini dilakukan untuk menganalisa data yang dikumpulkan dan kemudian akan digunakan untuk penelitian ini.

1. Dalam tahap pengumpulan data atau *scrapping*, peneliti mengambil data yang berasal dari Youtube yaitu data komentar pada sebuah video yang mengarah pada topik *Korean Wave*.
2. *Text Preprocessing*

Berikut tahapan dari Text Preprocessing yang digunakan:

(a) *Cleaning*

Pada tahap *Cleaning* yang dilakukan adalah menghapus karakter selain huruf, merubah teks menjadi huruf kecil (*case folding*), menghilangkan simbol atau karakter, *mention* atau *username*, menghapus *hashtag*, menghapus *URL* atau *link* dari setiap komentar.

(b) *Filtering*

Pada tahapan *filtering* dilakukan dengan membuang kata-kata yang tidak penting dari sebuah kalimat. Dan mengambil kata-kata yang

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

penting yang telah dipotong kata perkata.

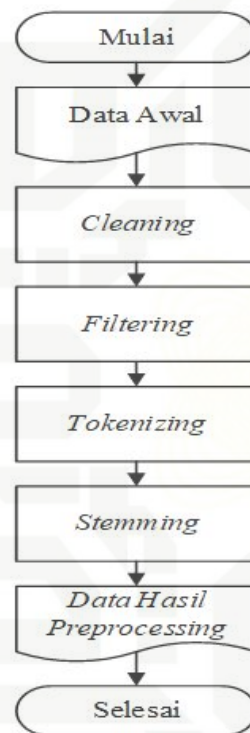
(c) *Tokenizing*

Tahap *tokenizing* adalah tahap memisahkan setiap teks menjadi setiap potongan kalimat dan kata berdasarkan spasi yang ditemukan.

(d) *Stemming*

Pada tahapan *Stemming* dilakukan dengan mengubah kata berimbuhan yang terdapat dalam dokumen menjadi kata dasar.

Alur pada proses preprocessing terhadap dokumen atau komentar dapat dilihat pada Gambar 3.2.



Gambar 3.2. Alur *Text Preprocessing*

3.3 Tahap Analisis dan Hasil

Adapun langkah-langkah yang dilakukan pada tahap analisa dan hasil adalah sebagai berikut:

1. Pelabelan

Pada tahap ini akan dilakukan pelabelan data yang telah diolah menjadi 3 label yaitu: positif, negatif, dan netral. Pelabelan data ini berguna untuk membuat hasil semakin lebih jelas.

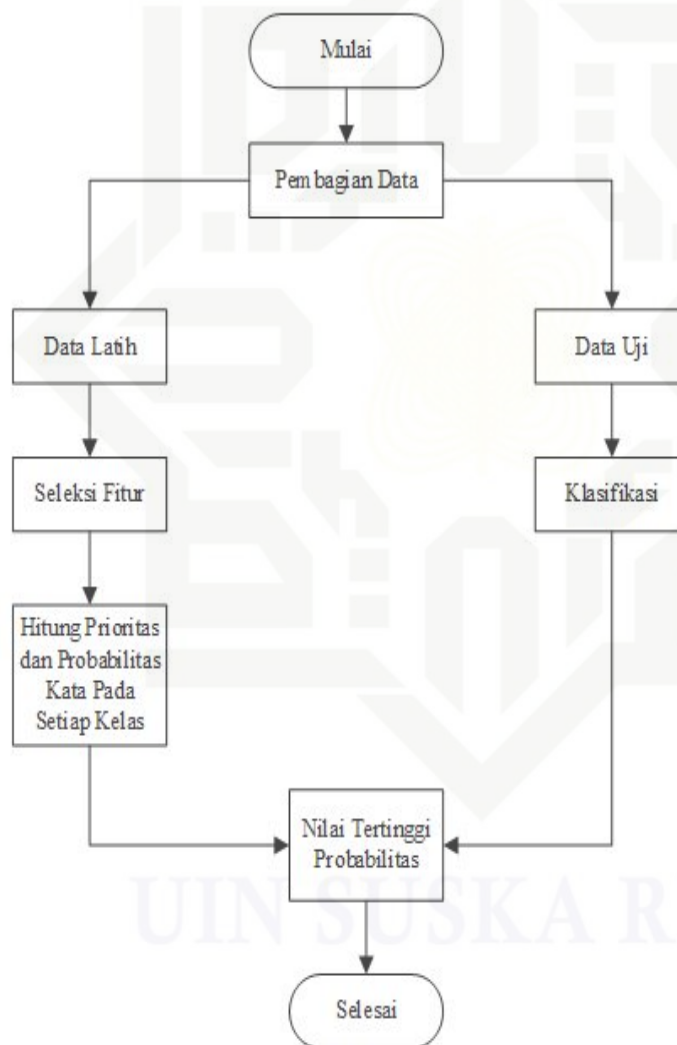
2. Pembobotan

Tahap pembobotan adalah tahap dimana setiap kata diberi bobot dengan menggunakan *Term Frequency* dan *Inverse Document Frequency* (TF- IDF).

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

3. Menentukan Data Latih dan Data Uji
Menentukan data latih dan data uji dengan menggunakan metode perhitungan *K-Fold Cross Validation*. Selanjutnya dilakukan menggunakan algoritma *Naïve Bayes Classifier* (NBC) setelah data tersebut diberi pembobotan.
4. Klasifikasi Data dengan *Naïve Bayes Classifier*
Pada tahap ini akan dilakukan pengklasifikasian menggunakan algoritma *Naïve Bayes Classifier* (NBC). Kemudian dilanjutkan dengan menghitung akurasi dari algoritma dan selanjutnya visualisasi data. Pada klasifikasi *Naïve Bayes Classifier* (NBC) terdapat beberapa langkah seperti pada Gambar 3.3.



Gambar 3.3. Alur Naive Bayes Classifier

3.4 Tahap Dokumentasi

Tahap terakhir adalah tahap melakukan dokumentasi terhadap seluruh kegiatan yang dilakukan pada Tugas Akhir ini mulai dari tahap perencanaan, pengumpulan data, dan proses analisa dan hasil, sampai tahap dokumentasi. dengan melakukan pembuatan laporan tugas akhir.

Hak Cipta Diindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

BAB 5 PENUTUP

5.1 Kesimpulan

Adapun beberapa kesimpulan yang dapat diambil dari penelitian Tugas Akhir ini adalah:

1. Pada data komentar Youtube tentang Korean Wave menggunakan algoritma *Naïve Bayes Classifier* mendapatkan hasil akurasi sebesar 62.75% dengan data *training* sebanyak 6300 data sedangkan data *testing* sebanyak 700 data.
2. Hasil sentimen terbanyak pada kelas positif dengan jumlah 671 komentar positif dengan kalimat yang sering muncul yaitu, “suka”, “kpop”, “lagu”, “musik” “korea”, “keren”. Dapat diartikan stigma yang berkembang di masyarakat mengenai Korean Wave didominasi oleh stigma positif.

5.2 Saran

Penelitian ini tentunya masih jauh dari kata sempurna dan tidak terlepas dari kekurangan, maka adapun saran dari peneliti adalah untuk penelitian selanjutnya melakukan percobaan menggunakan algoritma lainnya untuk melihat tingkat akurasi dari algoritma tersebut.

DAFTAR PUSTAKA

- Agusta, L., Kristen, U., dan Wacana, S. (2009). Perbandingan algoritma stemming porter dengan algoritma nazief & adriani untuk stemming dokumen teks bahasa indonesia. *Konferensi Nasional Sistem dan Informatika, 2009*, 196–201.
- Ahmed, K., dan Jesmin, T. (2014). Comparative analysis of data mining classification algorithms in type-2 diabetes prediction data using weka approach. *International Journal of Science and Engineering*, 7(2), 155–160.
- Ahn, J., Oh, S., dan Kim, H. (2013). Korean pop takes off! social media strategy of korean entertainment industry. Dalam *2013 10th international conference on service systems and service management* (hal. 774–777).
- Cayari, C. (2011). The youtube effect: How youtube has provided new ways to consume, create, and share music. *International journal of education & the arts*, 12(6), n6.
- Chandra, E., dkk. (2017). Youtube, citra media informasi interaktif atau media penyampaian aspirasi pribadi. *Jurnal Muara Ilmu Sosial, Humaniora, dan Seni*, 1(2), 406–417.
- Datareportal. (2021). *Digital 2021 : Indonesia*. Retrieved from <https://datareportal.com/reports/digital-2021-indonesia>
- Efendi, Z., dan Mustakim, M. (2017). Text mining classification sebagai rekomendasi dosen pembimbing tugas akhir program studi sistem informasi. Dalam *Seminar nasional teknologi informasi komunikasi dan industri* (hal. 235–242).
- Fayyad, U., Piatetsky-Shapiro, G., dan Smyth, P. (1996). From data mining to knowledge discovery in databases. *AI magazine*, 17(3), 37–37.
- Feldman, R., dan Sanger, J. (2007). *The text mining handbook: advanced approaches in analyzing unstructured data*. Cambridge university press.
- Hadna, M. S., Santosa, P. I., dan Winarno, W. W. (2016). Studi literatur tentang perbandingan metode untuk proses analisis sentimen di twitter. *Semin. Nas. Teknol. Inf. dan Komun*, 2016, 57–64.
- Juwita, S. H. (2018). Tingkat fanatisme penggemar k-pop dan kemampuan mengelola emosi pada komunitas exo-1 di kota yogyakarta. *Jurnal Riset Mahasiswa Bimbingan dan Konseling*, 4(7), 273–286.
- Lin, Y., Wang, X., dan Zhou, A. (2016). *Opinion spam detection. opinion analysis for online reviews, may*, 79–94.
- Ling, J., Kencana, I., dan Oka, T. B. (2014). Analisis sentimen menggunakan metode naïve bayes classifier dengan seleksi fitur chi square. *E-Jurnal*

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Matematika, 3(3), 92–99.

- Maulida, L. (2020). *Jumlah pengguna unik youtube di indonesia capai 93 juta*. Retrieved from <https://www.tek.id/tek/jumlah-pengguna-unik-youtube-di-indonesia-capai-93-juta-b1ZT79iPE>
- Mentari, N. D., Fauzi, M. A., dan Muflikhah, L. (2018). Analisis sentimen kurikulum 2013 pada sosial media twitter menggunakan metode k-nearest neighbor dan feature selection query expansion ranking. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer e-ISSN*, 2548, 964X.
- Misoch, S. (2014). Card stories on youtube: A new frame for online self-disclosure. *Media and Communication*, 2(1), 2–12.
- Mukherjee, S., Shaw, R., Haldar, N., dan Changdar, S. (2015). A survey of data mining applications and techniques. *vol*, 6, 4663–4666.
- Oktasari, L., Chrisnanto, Y. H., dan Yuniarti, R. (2016). Text mining dalam analisis sentimen asuransi menggunakan metode naive bayes classifier. Dalam *Prosiding seminar sains nasional dan teknologi* (Vol. 1).
- Rahmawati, A., Nikmah, N. L., Perwira, R. D. A., dan Rakhmawati, N. A. (2021). Analisis topik konten channel youtube k-pop indonesia menggunakan latent dirichlet allocation. *Teknologi: Jurnal Ilmiah Sistem Informasi*, 11(1), 16–25.
- Shim, D. (2006). Hybridity and the rise of korean popular culture in asia. *Media, culture & society*, 28(1), 25–44.
- Taheri, S., dan Mammadov, M. (2013). Learning the naive bayes classifier with optimization models. *International Journal of Applied Mathematics and Computer Science*, 23(4), 787–795.
- Talib, R., Hanif, M. K., Ayesha, S., dan Fatima, F. (2016). Text mining: techniques, applications and issues. *International Journal of Advanced Computer Science and Applications*, 7(11).
- Turban, A. d. L. T. P., E. (2005). *Introduction to data mining*. Pearson Education, Inc.
- Vijayarani, S., Ilamathi, M. J., Nithya, M., dkk. (2015). Preprocessing techniques for text mining-an overview. *International Journal of Computer Science & Communication Networks*, 5(1), 7–16.
- Widi, S. (2022). *Warga indonesia paling gemar tonton film korea lewat streaming*. Retrieved from <https://dataindonesia.id/digital/detail/warga-indonesia-paling-gemar-tonton-film-korea-lewat-streaming>

Hak Cipta Diindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

LAMPIRAN A

HASIL PELABELAN

Berikut ini merupakan hasil pelabelan data komentar video youtube dari beberapa *chanel* yang digunakan pada penelitian ini dapat dilihat pada Tabel A.1.

Tabel A.1. Hasil Pelabelan Data Komentar Youtube

No.	Teks	Label Pakar
1	Stuju belum lagi standar quot yg di pasangkan buat artis quot kpop	Netral
2	W suka lagunya chen exo suho exo touching and deep sih	Positif
3	iya sekarang tuh musik tuh kebanyakan lebih ke arah catchy yang penting enak ditelinga aku kadang kalo mo dengar yang lagu nyentuh tuh pake lagu ost drakor wkwk bagus bagus sih wkwk	Psitif
4	Aihhh br br mas hansol foto profil nya ganti gara quot iseng foto model waktu itu	Netral
5	Aku kalo ada group kpop muncul diyoutube ya gue buka streaming doang kalo bagus ya dilike kalo engga ya udah ga pernah ikut fanmet ke konsernya ataupun beli album ekwkwkk	Positif
6	Aku suka lagu lagunya btob	Positif
7	Aku tu dengar lagunya winner yg everyday gara gara lagu dan boyband itu aku suka exo ampe sekarang bucinnya sama do kyungsoo	Positif
8	Bts ka love myself magic shop intinya semua lagu dia menceritakan tentang hidup dia tentang bagaimana pentingnya dia mencintai diri sendiri aku termotifasi sama semua lagu bts aku jadi lebih cinta sama diri aku sendiri dan memandang beda dunia ternyata gaseburuk yang aku pikir positif	Positif
9	Suka heran sm yg suka k pop kok bisa anggota ny banyak bgt n bisa menghapal member ny kalau aku ngerasa anggota ny seragam suka susah bedain ny	
10	I hate kpop eh pas nonton twice auto suka njerrr itu awal gw suka kpop	Netral
11	Reaction milhya bang	Netral
12	Bener bgt setujuuu lagu lama tuh ga peduli penampilannya gimana yang penting lagunya enak dan ngena dihati banget persaingan tiap girl boy band masih ketat satu sama lain kalo jaman skrng kaya oke cakep lagunya jg yaa oke tp that s it ga ngena di hati kebanyakan lagunya dan persaingannya makin sempit	Netral
13	Jaman sekarang kalo nge fans kpop tuh pasti dikatain alay banyak yg gk suka sama kpopers juga gara fans nya yg pada bar sama toxic rata rata yg toxic pun fans yg masih anak alias bocah masya allah ketikannya ada yg keliatan bocah nya ada yg keliatan sok sok toxic maknnya jadi males	Positif
14	Aku terinspirasi utk pd jadi orang medok terimakasih	Netral
15	Sampai sekarang masih suka super junior	Positif

- Hak Cipta Diindungi Undang-Undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Table A.1. Hasil Pelabelan Data Komentar Youtube (Tabel Lanjutan...)

No.	Teks	Label Pakar
16	Kalo aku sih setelah ne bubar kecewa banget idolaku harus disband se- menjak itu aku udah jarang dengar lagu kpop	Netral
18	Yang penting ngelakuin yang berfaedah aja dah	Positif
19	Klo gua sih emg suka kpop tp g suka banget soalnya buat apa beli banyak koleksi kpop pdhl kpop g terkenal selamanya v	Netral
20	Mas hansol sama nih kayak saya awal suka k pop dan karena big bang the only grup at that time that move my heart wkwkwk di antara skian boygrup yg waktu itu pada pakai konsep pretty boy hingga bkin image cowo k pop banci dan waktu itu k pop ers minoritas sering dbully gara image banci amp oplas padahal mreka gtw aja klo big bang jauh dari kesan itu and bener bgt stelah kualitas musik k pop menurun menurut saya jd lah saya berhenti mengikuti k pop tapi tetap vip di hati secara lagu bigbang sudah mengisi hari sma amp kuliah saya sbg vip saya suka bigbang karena musiknya ya jadi ngambilnya yg positive nya aja yg jelek nya jgn ditiru toh gd jg dl bilang gitu	Netral
21	Senang kpop dari ost nya drakor aja lebih ngena aja maknanya krn dibarengi moviesnya	Positif
22	Sudah coba dengarkan bts mereka benar benar membagikan perasaan dan hidup mereka lewat musik dan lirik	Positif
23	Kpopers hidup dalam hayalan	Netral
24	Bts lagunya dalem mas hansool	Netral
25	Setuju bgt thats why i always luv bigbang selalu menyentoeehhh	Positif
26	Aku dlu suka anime sekarang suka k drama	Netral
27	Yg bikin aku suka kpop sejak tau sama beast skg jadi highlight abis itu cinta banget sama bts pas awal debut army belum sebanyak skg mer- ambah ke drakor akhirnya stop ngikutin setelah punya anak dan pindah haluan ke pageant	Positif
28	Gw malah salfok ama jidat muu mengkilat gara gara ada lampu terlalu terang sih	Netral
29	Dulu aku juga mikir gitu tapi ini gara aku ga sengaja dengar ceramah- nya namjoon rm bts kelar sudah karena setiap lagu yg mereka buat pun- ya deep meaning haha br coba deh mas tapi harus open minded tidak memaksa karna sampe sekarangpun masih cuma suka bts gabisa suka	Netral
30	Quot misalkan subscriberku juta quot dan skrg pas video ini muncul di homeku subscriber sam udah juta	Netral
31	Good	Netral
32	While lagu wibu jijik di dengar cewe tapi memiliki arti dan makna yg mendalam	Positif

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Table A.1. Hasil Pelabelan Data Komentar Youtube (Tabel Lanjutan...)

No.	Teks	Label Pakar
33	Gw pas kelas dulu juga suka kpop terus girlband boyband favoritku red velvet blackpink sama bts br klo skrg sih skrg gw lebih suka anime br alasannya gini jdi kan dulu gw feminim terus suka kpop terus gw dibully sama temen gw cuma gara gara gw suka kpop karena gw nggak tahan gw jadi tomboy dan jadi wibu biar gak dibully lagi br br br br br maapkeun cerita gw cukup gak jelas v	Netral
34	Aq mah tetep ne	Netral
35	Mas skrg subscribes mu jt yeeii	Netral
36	Iya aku juga udah gak segitu ngikutin kpop banget hmm dulu sih ya aku suka tvxq doang hehehe ntah sekarang karena faktor usia yang udah mau kepala jadi otomatis gak ngikutin lagi tapi masih suka pcy	Netral
37	Enak ditelinga danyg menyentuh hati untuk sekarang tuhh day hehe	Positif
38	Btw suka banget dengar mas hansol ngomong suka logatnya aku yg asli jawa ajah gak gitu logatnya pegel malah pas nyoba kayak mas hansol kerenlah pokoknya btw ini pake akun suamiku	Positif
39	Ternyata kita sama dong tdak berminat kpop	Netral
40	Iya sih aku suka kpop tapi mengagumi bukan berarti tidak cinta tanah air aku hanya mau mencari makna dari lagu lagu luar karena di indonesia sendiri udah jarang banget lagu yang ada maknanya t t	Positif
41	Berlaluanya waktu emang selera itu berubah baik dr music makanan dan selera baju soalnya q jg mengalami itu makannan yg dulu q g suka pare eh skrg sdh ibuk ibuk q suka makan pare q bisa menikmati rasa pait pare itu q suka jahe dan rempah rempah pdhl waktu q g sukaa rempah dan q rasa oh iya ya manusia itu berubah	Positif
42	Mon maaf aku ga fokus ke mas hansol tapi ke komen netijen	Netral
43	Lulus sma kita seumuran haha	Netral
44	Dri dlu smpi skrang aq ttap suka metallica	Netral
45	Klo ak sih semenjak kenal bts jadi lebih fokus ke belajar dan pengen ktmu idola pake uang hasil kerja nnti makanya apa apa sekarang lebih merencanakan lagi sllu bikin agenda buat besok nanti banyak lagu bts yang menginspirasi dan klo dipahami maknanya itu nusuk gt lho jadi nyess klo dengar tau artinya	Positif
46	Hansol tell me tell me br me arrrggghh le jong suk	Netral
47	Bahas tentang ini sekarang kebanyakan war sedih cuma banyak gd aja udah dikatain	Negatif
48	i]Intinya udah mainstream yo musiknya br kayak ga terlalu bermakna liriknya br br meski ga semua sih tapi rata emang hambare	Netral
49	Setuju	Positif
51	Era terbaik kpop bukan di era billboard amerika	Positif
52	Jadi jawaban untuk keluhan ibu ibu itu ada di video mana guys	Negatif

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Table A.1. Hasil Pelabelan Data Komentar Youtube (Tabel Lanjutan...)

No.	Teks	Label Pakar
53	Ga tertarik dari dulu sih ya tau sih tapi ya seperti mas hansol bilang ga sampe yg menggugah hati saya cuma tertarik dg perjuangan mereka sampe jadi besar musik tetep dari dlu metal walupun banyak orang yg bilang musiknya ga jelas tapi saya suka korea dalam hal fashion	Positif
54	Aku suka e lagu lagu e park hyo shin mergo suarae uenak poll	Positif
55	Mr korea reomit saya suka lagu dari kim boon soo adakah itu termasuk suka kpop ke harap balas pertanyaan ini	Positif
56	Klo jaman sma kuliah suka kpop tp cuma yg lg trending aja siy skrng sukanya kdrama itupun yg saeguk n fantasy	Netral
57	Try j pop	Netral
58	Ternyata ada juga yg berpikiran sama kayak aku aku suka kpop udh lama dan awalnya juga tau dari kakak karna sering buka lagu nya entah kenapa makin kesini lagu kpop ny itu udh enggak pernah aku dengar tp aku masih sering buka lagu kpop yang lama ya karna ngerasa gak cocok aja ditelinga ini pendapat aku ya tapi aku masih suka kpop kok	Netral
59	Dhlh gw kangen smash cherrybelle xo ix ama jkt grup aja lh kangen bat	Netral
60	Soal lagu itu selera sih tapi semisal kalian suka sama salah satu idol yasudah cukup suka aja masalah fandomnya toxic suka war dll ya itu nggakbisa kalian kontrol toh disetiap fandom pasti ada saja yang begitu tapi kita juga enggak berhak mengeneralisir kalau fandom a toxic fandom b suka war etc suka ya suka aja tanpa mikir fandom yang bla bla bla suka yang	Negatif
61	Mas omah e ngdi rek korea iso medok mas wkwk	Netral
62	Aq under estimate sama cowo korea karna liat mereka lemah gemulai pake anting panjang pake lipstik ilfeel gimanaa gitu tapi ko mas hansol maskulin ya apa cuma artis nya yg feminism	Negatif
63	Pen berhenti suka kpop tpi gak bisaa loh mas	Netral
64	Oce setuju mz brrooww aqw yoo ga terlalu suxa xpop	Netral
65	Apa untung nya ngefans sama artis artis gitu apa artis itu bisa memberikan anda hidup ada juga fans nya mimpi mimpi mau jadi suami atau istri nya gitu namanya juga terlalu halu jadi boleh aja ngefans tapi jangan sampe ketinggian ya guys	Netral
66	Aku kpopies sejak liat mv the boys snsd im sone yes but sejak member beda agensi aku jadi rada kurang ke kpop tapi tiba tiba twice chaeyoung meruntuhkan tembok itu jadi niat mau pensiun dri perkpopan batal sudahh	Negatif

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Table A.1. Hasil Pelabelan Data Komentar Youtube (Tabel Lanjutan...)

No.	Teks	Label Pakar
67	Drakornya juga skrg pada kaga warass sumpah aku terkejoedd ketika melihat ada drakor yg berbau gay judulnya fake love dan drama itu ada dibintangi ama penyanyi k pop sama aktor yg aku idolain rata quot komen an nya setuju quot aja ada juga yg enggak soo kalo kalian bilang begini what s wrong with being a gay lesbian the answer is itu totally salah kalo kita belajar sosiologi itu termasuk perilaku menyimpang gak cukup oke jadi jawaban lagii tuhan menciptakan cewek dan cowok untuk hidup secara berdampingan jadi cowok itu sudah ada pendamping nya yaitu cewek dan juga sebaliknya ku harap bolo quot disini gak ada yg gitu benar quot minta maaf kalo ada kata yg ga berkenan yaa	Netral
68	Korea vs jepang br bias vs waifu	Negatif
69	Kata kpopers lets kill this loveee ay ayayayaya aku rapooopoo rapooopooo	Netral
70	Dulu directioners br br terkenal banget dong kalau directioners itu sering war sama exo l dan army br hingga gua nemu vid one direction vs exo br dan itu gua makin suka d amp exo br padahal dulu anti k pop cuma suka k drama doang	Negatif
71	Btob	Netral
72	Samaaaaa suka big bang karena lagu itu juga	Positif
73	Pengen banget berhenti ngikutin kpop br demi masadepan sukses	Positif
74	Saya awal juga gak suka kpop tapi suka sekarang itupun hanya exo awalnya gara nonton drama korea yg ada idolnya trus saya search koq aktingnya bagus ya dan baru tau mereka tuh idol dan group pula trus dengar lagu mereka dan ternyata suara mereka tuh bagus dan bukan modal tampang aja	Positif
75	Salfok pas sampean ngomong pendapatan wkwkwk br nek aku dulu suka kpop as tahun iku pas ft island lagi joyo joyone karo superjunior sing happines berlanjut menggila sampe lah terusan e aku dadi melu terhipnotis dadi fangirlingan trus lama kelamaan tak pikir pikir kok gak onok faedah e yo dadi fangirlingan akhire aku dadi biasa ae yo tetep suka sewajar e saiki makin dewasa makin realistis aku nek fangirlingan terus eman eman uangnya bisa digawe tuku jajan dan ditabung ae wes lagian nek awakdewe dadi fans trus artise loh yo gak ngerti awakdewe awakdewe histeris karepe dewe artise lempeng nang panggonan e kono artise makin subur awakdewe malah boncos bokek dsb	Negatif
76	Iyooo masss hansollll koyoo q sm kamuu mass kecanduaann nontonnn setiap hari jdi rutinans q nonton chanell mas hansoll br br trs berkarya membagikan konten quot yg menghibur bermanfaat sukseess wat mas hansollllll br saranghae mas hansol	Positif
77	Gua suka kpop dari thn saat era baru bigbang fantastix baby dan bigbang udah mulai out of the box tahun ditahun tsb	Negatif
78	Tambah ndutt	Netral

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Table A.1. Hasil Pelabelan Data Komentar Youtube (Tabel Lanjutan...)

No.	Teks	Label Pakar
79	Apa lagu instrument di drakor terhitung k pop gua suka tuh tapi kalo kpop format boyband atau girlband nggak begitu suka itu elektronik semua kebanyakan sih	Netral
80	Kalo aku itu gak suka k pop bukan benci itu karena fans nya itu terlalu berlebihan yang buat gua jadi yg awal nya pengen coba liat jadi males	Negatif
81	Btull	Netral
82	Aku kpopers dr jaman sklh mpe skrg udah emak br yg penting gak smpe lupa waktu aja kewajiban ibu rumah tangga no kpop ya menyusul	Netral
83	Mngapa nama nama orang korea tidak jauh dari kata kim park lee dan bagaimana membedakan jika nama itu nama orang wanita atau pria misal kalau di indonesia sarmo itu laki dan bila sarmi itu pasti cewek	Negatif
84	Aku tetep sm stan sih apalagi exo wkwk gaberubah meskipun sekarang banyak yg baru	Positif
85	Bisa lancar gitu ya lagunya	Netral
86	Cm mo tanya kenapa artis korea itu kurg terbuka soal kehidupan probadi mereka ya bahkan banyak dr mereka yang ga punya akun ig knp apa karena fans mereka yang serem abis	Positif
87	Sama aku juga suka kpop jama dulu ketimbang sekarang	Netral
88	Aku juga ngerasa gtu mas hansol ga cuma kpop	Netral
89	Aku jg berhenti suka k pop dari januari yaa aku juga punya alasan yg aneh sih v	Negatif
90	Nonton chanel ini adsny sama terus	Netral
91	Oppa lulus sma wew seketika aq merasa tua bgt aq lulus sma and aq udh pny anak seketika jiwa tua ku meluap	Netral
92	Aku setuju banyak lagi perkara yg penting	Positif
93	Saya berhenti jadi fans kpopers amp berhenti suka kpop br saya berlalih ke musik Hollywood	Netral
94	Gue dulunya enek banget sama kpop tapi semenjak kenal exo hmm jadi bucin	Netral
95	Ko setuju ya aku jg smenjak lulus sma udah ga bgtu suka kpop yg di-puter cuma lagu yg dulu sbml th an	Positif
96	Setujuuuu kang	Netral
97	Setuju lagy quot jmn dulu lbih nyampe	Positif
98	Aku berhenti suka kpop karna selalu disindir sama guru ku dan lama aku sadar oh iya guruku bilang kayak gini karna kasian sama aku karna kalo aku jadi kpopers setiap hari aku di bully tmn aku br br gtu cerita nya	Negatif
99	Itu pun yang aku rasakan terhadap indonesia dalam hal musik dan dramanya br sebenarnya semua musik baik itu western k pop jndonesian musik dan lain nya zaman sekarang tuh sulit menemukan sesuatu yg menyentuh	Positif

- Hak Cipta Diindungi Undang-Undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
 2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Table A.1. Hasil Pelabelan Data Komentar Youtube (Tabel Lanjutan...)

No.	Teks	Label Pakar
100	Sama permikiran nih waktu tahun aku tuh emang kpopers tegar tiap hari ku itu dipenuhi dengan k pop k pop k pop panutan ku ketika itu group super junior sampe ditahun aku jadi histeris sama exo group baru yang debut ketika itu nah diujung tahun gak tau kenapa perlahan lahan aku mulai slow sampe sekarang ni aku cuman nonton running man kalo dulu asal rancangan televisi yang ada k popnya semua aku nonton skrg yah rm doang kalo bahas group sih cuman group exo dan nue st aja ni yang sempat jadi group favorit terakhirku sebelum aku berhenti jadi kpopers tegar nah bagiku setelah eranya exo gak ada lagi nih group yang benar bikin aku jatuh cinta kayak suju. ...	Positif
6991	Jangan berhenti dulu coba react kpop secret number o masak yg lagi viral karena idol kpop pertama kali dari indonesia tidak di react sih debut sebulan lalu sudah juta views lho sssst anak dari indonesia itu bukan kaleng kaleng dia keturunan ningrat berdarah biru dari kerajaan di bali dari kasta kesatria ayo react lagi heboh hebohnya secret number adalah rookie monster	Positif
6992	Aku suka kpop karna exo	Positif
6993	Gua sekarang jadi wibu	Positif
6994	Dulu kpop lagu lagunya lagu yang bisa dinyai in dan di acara ajang bakat cocok untuk di bawa in di panggung sedangkan skarang lagu nya lagu yang rata kalimat nya di ulang dan harus easy listening dan ber ciri khas intinya kalo dulu itu fokus nya di nyanyian dan skarang kpop fokus nya di musik banget contoh nya wonhae manhi manhi manhi manhi gitu terus di ulang itu kan easy listeing bgt dan gampang di inget untuk dluar korea tpi kualitas untuk di nyanyi in aneh tapi pasar skarang yang laku gitu beda ama haru haru atau lies nya bigbang yang di nyanyi in mas hanso lagunya emang bener lagu cocok dan enak di nyanyi in dan lebih menyentuh jadi kpop generasi alias generasi eranya bigbang kelar emang pasar nya dan fokus lagunya dah beda lebih ke arah yg easy listening atau di ingat bukan yang enak dinyanyiin sebagai penyanyi	Positif
6995	Siapa exo l vip atmy blink like kali aowkawkowkwkwkwkwkwkwkwkwoowkwowkwoakowkskwowkwokwowlpwo	Positif
6996	Yang laki laki suka nonton drakor buat ngisi waktu sini ngumpul	Netral
6997	Lagu bigbang the best	Positif
6998	Wah iya bener juga ya pendapatnya mas hansol kalau kpop zaman now lagunya kurang ngena di hati atau bahkan ga sama sekali even tho lagunya enak didengar vibesnya beda aja sama yang kpop lama especially kpop nd generation mind blowing parah nih mas hansol	Positif
6999	Baru liat inii haha sama kok aku juga sejak kuliah udah ga update sama k pop karena kayak krg engage aja terakhir paling aja suka sama yg jadul suju big bang pm dbks	Positif

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Table A.1. Hasil Pelabelan Data Komentar Youtube (Tabel Lanjutan...)

No.	Teks	Label Pakar
7000	Aku sampai tergilgila gila sama bts	Positif



- Hak Cipta Dilindungi Undang-Undang**
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
 2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Hak Cipta Diindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

LAMPIRAN B

DAFTAR *STOPWORDS*

Berikut ini merupakan daftar kata stopword yang digunakan pada penelitian ini dapat dilihat pada Tabel B.1.

Tabel B.1. *Stopword*

No.	Stopword	No.	Stopword	No.	Stopword
1	Ada	65	Belakang	129	Bukannya
2	Adalah	66	Belakang	130	Bulan
3	Adanya	67	Belum	131	Bung
4	Adapun	68	Belumlah	132	Cara
5	Agak	69	Benar	133	Caranya
6	Agaknya	70	Benarkah	134	Cukup
7	Agar	71	Benarlah	135	Cukupkah
9	Akankah	73	Berakhir	136	Cukuplah
10	Akhir	74	Berakhirlah	137	Cuma
12	Akhirnya	76	Berapa	138	Dahulu
13	Aku	77	Berapakah	139	Dalam
14	Akulah	78	Berapalah	140	Dan
15	Amat	79	Berapapun	141	Dapat
16	Amatlah	80	Berarti	142	Dari
17	Anda	81	Berawal	143	Daripada
18	Andalah	82	Berbagai	144	Datang
19	Antar	83	Berdatangan	145	Dekat
20	Antara	84	Beri	146	Demi
21	Antaranya	85	Berikan	147	Demikian
22	Apa	86	Berikut	148	Demikianlah
23	Apaan	87	Berikutnya	149	Dengan
24	Apabila	88	Berjumlah	150	Depan
25	Apakah	89	Berkali-kali	151	Di
26	Apalagi	90	Berkata	152	Dia
27	Apatah	91	Berkehendak	153	Diakhiri
28	Artinya	92	Berkeinginan	154	Diakhirnya
29	Asal	93	Berkenan	155	Dialah
30	Asalkan	94	Berlainan	156	Diantara
31	Atas	95	Berlalu	157	Diantaranya
32	Atau	96	Berlangsung	158	Diberi
33	Ataukah	97	Berlebihan	159	Diberikan
34	Ataupun	98	Bermacam	160	Diberikannya
35	Awal	99	Bermacam-macam	161	Dibuat
36	Awalnya	100	Bermaksud	162	Dibuatnya
37	Bagai	101	Bermula	163	Didapat

Table B.1. *Stopword* (Tabel Lanjutan...)

38	Bagaikan	102	Bersama	164	Didatangkan
39	Bagaimana	103	Bersama-sama	165	Digunakan
40	Bagaimanakah	104	Bersiap	166	Diibarat
41	Bagaimanapun	105	Bersiap-siap	167	Diibaratkannya
42	Bagi	106	Bertanya	168	Diingat
43	Bagian	107	Bertanya-tanya	169	Diingatkan
44	Bahkan	108	Berturut	170	Diinginkan
45	Bahwa	109	Berturut-turut	171	Dijawab
46	Bahwasanya	110	Bertutur	172	Dijelaskan
47	Baik	111	Berujar	173	Dijelaskannya
48	Bakal	112	Berupa	174	Dikarenakan
49	Bakalan	113	Besar	175	Dikatakan
50	Balik	114	Betul	176	Dikatakannya
51	Banyak	115	Betulkah	177	Dikerjakan
52	Bapak	116	Biasa	178	Diketahui
53	Baru	117	Biasanya
54	Bawah	118	Bila	748	Wah
55	Beberapa	119	Bilakah	749	Wahai
56	Begini	120	Bisa	750	Waktu
57	Beginian	121	Bisakah	751	Waktunya
58	Beginitah	122	Boleh	752	Walau
59	Beginitulah	123	Bolehkah	753	Walaupun
60	Begitu	124	Bolehlah	754	Wong
61	Begitukah	125	Buat	755	Yaitu
62	Begitulah	126	Bukan	756	Yakin
63	Begitupun	127	Bukankah	757	Yakni
64	Bekerja	128	Bukanlah	758	Yang

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

LAMPIRAN C

SURAT PERNTATAAN

SURAT PERNYATAAN PAKAR BAHASA

Saya yang bertanda tangan dibawah ini:

Nama : Nuratika, S.Hum., M.Pd
Tempat/Tanggal Lahir : Sebaling/ 02 Desember 1988
Pekerjaan : Dosen Pendidikan Bahasa dan Sastra Indonesia
STKIP Rokania Rokan Hulu Riau
Alamat : Ujungbatu Rokan Hulu Riau

Menyatakan dengan sesungguhnya bahwa benar memvalidasi untuk pelabelan kata yang bersifat Positif, Netral dan Negatif pada data komentar youtube sebanyak 7.000 data dalam Tugas Akhir dari:

Nama : Irina
NIM : 11753200071
Jurusan : Sistem Informasi
Judul Tugas Akhir : Penerapan *Naïve Bayes Classifier* Untuk Analisis Sentimen Terhadap Pengaruh *Korean Wave* Di Indonesia Pada Youtube

Demikian surat pernyataan ini saya buat dengan sesungguhnya tanpa paksaan dari pihak manapun. Atas pengertiannya saya ucapkan terima kasih.

Ujungbatu, 25 Januari 2023
Yang membuat Pernyataan



Nuratika, S.Hum., M.Pd.

DAFTAR RIWAYAT HIDUP



© Hak cipta milik UIN Suska Riau

Irina Lahir di Seberang Tembilahan, Indragiri Hilir, Riau pada tanggal 11 Februari 1999 dari Bapak Ali dan Ibu Ika-naria. Penulis merupakan anak kedua dari dua bersaudara. Riwayat pendidikan yang dilalui penulis dimulai Sekolah Dasar Negeri 016 Sungai Perak, Tembilahan pada Tahun 2006-2012, dilanjutkan di SMP Negeri 004 Sungai Perak, Tembilahan, Indragiri Hilir pada Tahun 2012-2014. Setelah lulus SMP, penulis melanjutkan pendidikan di Sekolah Menengah Atas di SMA Negeri 02 Tembilahan, Indragiri Hilir pada Tahun 2014-2017. Kemudian penulis melanjutkan studi S1 pada Tahun 2017 di Universitas Islam Negeri Sultan Syarif Kasim Riau, Pekanbaru dengan Program Studi Sistem Informasi Fakultas Sains dan Teknologi. Selama menjalani masa studi studinya, penulis pernah melaksanakan kerja praktek di Kementerian Agama Kota Pekanbaru. Penulis juga mengikuti pengabdian Kuliah Kerja Nyata (KKN) di Desa Sungai Perak, Kabupaten Indragiri Hilir. Penulis merupakan salah satu anggota aktif pada organisasi kampus yaitu *Puzzle Research Data Technology*. Serta melakukan Penelitian Tugas Akhir dengan Judul **“Penerapan Naïve Bayes Classifier Untuk Analisis Sentimen Terhadap Pengaruh Korean Wave Di Indonesia Pada Youtube”**. Untuk menjalin komunikasi dengan penulis baik dikampus maupun diluar kampus dapat menghubungi email irina78822016@gmail.com.

UIN SUSKA RIAU

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
 2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.
- UIN SUSKA RIAU