

**PENERAPAN *NAIVE BAYES CLASSIFIER* UNTUK ANALISIS
SENTIMEN MASYARAKAT TENTANG PENANGGULANGAN
BENCANA ALAM DI INDONESIA**

TUGAS AKHIR

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

Oleh:

NURHALIMAH

11753200030



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.

**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 MASYARAKAT TENTANG PENANGGULANGAN BENCANA ALAM DI INDONESIA****TUGAS AKHIR**

Oleh:

NURHALIMAH**11753200030**

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

UIN SUSKA RIAU



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 mengumutikan 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 MASYARAKAT TENTANG PENANGGULANGAN BENCANA ALAM DI INDONESIA

TUGAS AKHIR

Oleh:

NURHALIMAH

11753200030

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



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:

© Hak cipta milik UIN Suska Riau
Nama : Nurhalimah
NIM : 11753200030
Tempat/Tgl. Lahir : Simpang Tonang / 14 Maret 1999
Fakultas/Pascasarjana : Sains dan Teknologi
Prodi : Sistem Informasi
Judul Disertasi/Thesis/Skripsi/Karya Ilmiah lainnya*:
PENERAPAN NAIVE BAYES CLASSIFIER UNTUK ANALISIS
SENTIMEN MASYARAKAT TENTANG PENANGGULANGAN
BENCANA ALAM DI INDONESIA

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


NURHALIMAH
NIM : 11753200030

*pilih salah satu sesuai jenis karya tulis



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.

State Islamic University of Sultan Syarif Kasim Riau

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 kedua orang tuaku tercinta Ayahanda Sahril dan Ibunda Ratna Dewita yang telah merawat dan membesarkan saya, memenuhi segala yang saya butuhkan hingga saat ini, yang selalu sabar menemani, menanti, memberi support dan mendoakan perjalanan anaknya untuk menjadi seorang Sarjana yang ilmunya dapat bermanfaat bagi orang-orang di Sekitar. Semoga Allah melimpahkan keberkahan, kesehatan dan umur yang panjang, walau jasamu tak mampu terbalaskan, aku berharap selalu bisa membahagiakanmu. Selanjutnya untuk kakak-kakak dan abang tersayang Alia Netti, Febrian Doni dan Nurazizah serta ipar tersayang Yuliarti yang sudah melimpahkan kasih sayang dan menjaga adik bungsunya sampai sekarang serta membantu banyak dalam finansial. Semoga Allah melimpahkan keberkahan, kesehatan, umur yang panjang dan kelancaran rezeki. Serta Keluarga besar yang selalu memberikan doa, bantuan, dan semangat kenapa ini harus diselesaikan. Semoga Allah SWT selalu memberikan kita kemudahan dalam menjalankan setiap urusan, dilimpahkan taufik dan hidayahnya, Aamiin. Salam sayang dan rindu selalu untuk Keluargaku, dimanapun berada.

KATA PENGANTAR

Alhamdulillah rabbil 'alamin, puji syukur penulis ucapkan kehadiran Allah SWT yang telah melimpahkan rahmat dan karunia-Nya sehingga penulis dapat menyelesaikan Tugas Akhir dengan judul “Penerapan *Naive Bayes Classifier* untuk Analisis Sentimen Masyarakat Tentang Penanggulangan Bencana Alam di Indonesia”. Penulisan Tugas Akhir ini dimaksudkan untuk memenuhi salah satu syarat dalam rangka menyelesaikan studi Strata 1 (S1) di Universitas Islam Negeri Sultan Syarif Kasim Riau. Tak lupa shalawat beserta salam selalu tercurahkan kepada Nabi Muhammad SAW dengan mengucapkan “*Allahummasolli ‘alamuhammad, wa’alaalimuhammad*” yang telah menjadi suri tauladan yang baik bagi kita semua.

Dalam penyusunan dan penyelesaian Tugas Akhir ini, penulis menyadari bahwa Tugas Akhir ini tidak akan terwujud dengan baik tanpa adanya bantuan dari semua pihak, untuk itu penulis menyampaikan ucapan terima kasih banyak kepada:

1. Bapak Prof. Dr. Hairunas, M.Ag selaku Rektor Universitas Islam Negeri Sultan Syarif Kasim Riau.
2. Bapak Dr. Hartono, M.Pd selaku Dekan Fakultas Sains dan Teknologi.
3. Bapak Eki Saputra, S.Kom., M.Kom selaku Ketua Program Studi Sistem Informasi.
4. Ibu Siti Monalisa, ST., M.Kom selaku Sekretaris Program Studi Sistem Informasi.
5. Bapak M.Afdal ST., M.Kom selaku Pembimbing Akademik selama saya berkuliah hingga Tugas Akhir ini selesai.
6. Ibu Dr. Rice Novita, S.Kom., M.Kom sebagai Dosen Pembimbing Tugas Akhir saya.
7. Bapak Tengku Khairil Ahsyar, S.Kom., M.Kom selaku Ketua Sidang Tugas Akhir saya dan selaku Kepala Laboratorium Program Studi Sistem Informasi.
8. Bapak Syaifullah, SE., M.Sc selaku Penguji I Sidang Tugas Akhir saya
9. Ibu Fitriani Muttakin, M.Sc selaku Penguji II Sidang Tugas Akhir saya.
10. Kepada Seluruh Dosen dan Staff Program Studi Sistem Informasi Fakultas Sains dan Teknologi.
11. Keluarga Tercinta, Bapak, Umak, Uni Alianetti, Abang Febrian Doni, dan Uni Nurazizah Kak Yuliarti yang merupakan keluarga besar tercinta, penulis sangat mengucapkan terimakasih atas semua doa’a dan dukungannya dalam bentuk apapun, yang menjadi motivasi dan inspirasi untuk terus melangkah.
12. Untuk sahabat dan teman-teman Sistem Informasi Kelas D 2017 serta Sis-



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.

tem Informasi 2017 yang telah banyak membantu dan memberikan masukan dan semangat serta motivasi.

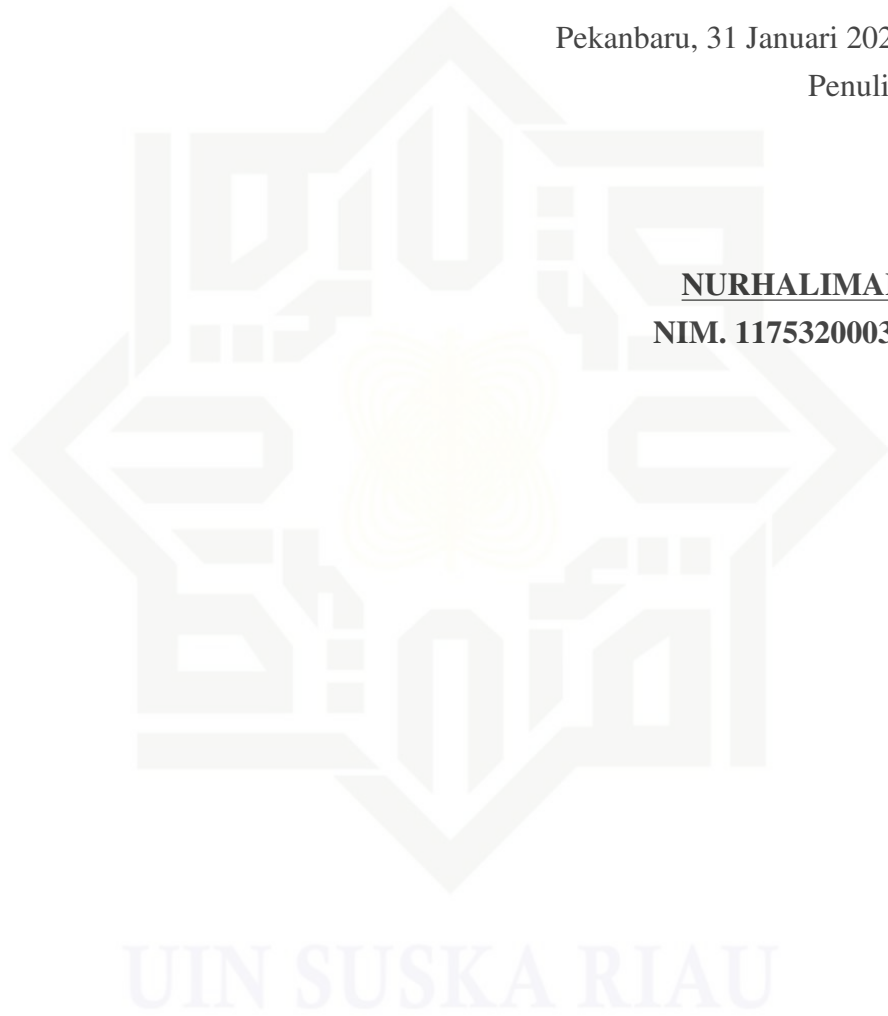
13. Untuk seluruh keluarga besar Program Studi Sistem Informasi yang telah membantu dan memberikan masukan positif kepada penulis.

Semoga dengan segala jerih payah dan dorongan yang telah disumbangkan, bernilai sebagai amal ibadah di sisi Allah SWT, Amin. Tugas Akhir ini masih jauh dari kesempurnaan, untuk itu diharapkan kritik dan saran yang sangat membangun dari pembaca dan dapat disampaikan ke email nurhalimah1499@gmail.com. Atas bantuannya saya ucapkan terima kasih. *Wassalamu'alaikum Wr. Wb.*

Pekanbaru, 31 Januari 2023

Penulis,

NURHALIMAH
NIM. 11753200030





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 MASYARAKAT TENTANG PENANGGULANGAN BENCANA ALAM DI INDONESIA

NURHALIMAH
NIM: 11753200030

Tanggal Sidang: 29 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

Perkembangan teknologi yang berjalan dengan pesat memberikan pengaruh yang besar terhadap kehidupan masyarakat, salah satunya dalam hal penggunaan aplikasi sosial media. *Youtube* merupakan salah satu layanan dari *Google* yang memfasilitasi penggunaannya untuk mengupload dan mengunduh video dan bisa diakses oleh pengguna lain secara gratis. Bisa dikatakan *Youtube* adalah *database* video yang paling populer di dunia internet, bahkan yang paling lengkap dan variatif. Adapun reaksi masyarakat dapat digali melalui komentar yang ditinggalkan pada kolom komentar setiap video. Bencana alam sering kita lihat di media sosial karena Indonesia termasuk negara yang tingkat bencana alam yang tinggi. Dari peristiwa ini memicu masyarakat dalam menyampaikan opini dalam bentuk komentar berupa ungkapan rasa sedih, peduli atau keluhan. Kumpulan data tersebut dapat diolah dengan menggunakan konsep *Text Mining*. *Text Mining* dilakukan dengan ditambahkan sebuah algoritma klasifikasi yaitu *Naive Bayes Classifier* (NBC). Penelitian ini bertujuan untuk menguji akurasi algoritma *Naive Bayes* untuk analisis sentimen opini masyarakat dari data komentar video *Youtube* tentang bencana alam, dan menemukan informasi dan sentimen terkait komentar bencana alam di *youtube*. Data yang didapatkan dari komentar sebanyak 7608 dengan data *training* sebanyak 6843 data *training* dan 761 data *testing*. Hasil penelitian ini menghasilkan akurasi sebesar 55,71%. Hasil dari sentimen negatif 585, sentimen netral 168, dan sentimen positif 8.

Kata Kunci: Bencana Alam, *Naive Bayes Classifier*, *Youtube*.



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 MASYARAKAT TENTANG PENANGGULANGAN BENCANA ALAM DI INDONESIA

NURHALIMAH
NIM: 11753200030

Date of Final Exam: January 29th 2023
Graduation Period:

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

ABSTRACT

The rapid development of technology has had a major impact on people's lives, one of which is the use of social media applications. Youtube is a service from Google that facilitates its users to upload and download videos and can be accessed by other users for free. "You could say Youtube is the most popular video database on the internet, even the most complete and varied. The public reaction can be explored through the comments left in the comments column for each video. We often see natural disasters in social media because Indonesia is a country with a high level of natural disasters. This event triggered the community to express opinions in the form of comments in the form of expressions of sadness, concern or complaints. The data set can be processed using the concept of Text Mining. Text Mining is done by adding a classification algorithm, namely the Naïve Bayes Classifier (NBC). This study aims to test the accuracy of the Naïve Bayes algorithm for sentiment analysis of public opinion from data on YouTube video comments about natural disasters, and to find information and sentiment related to natural disaster comments on YouTube. The data obtained from the comments is 7608 with training data of 6843 training data and 761 testing data. the results of this study resulted in an accuracy of 55.71 %. The result is 585 negative sentiments, 168 neutral sentiments, and 8 positive sentiments.

Keywords: *Natural Disasters, Naive 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	4
1.3 Batasan Masalah	4
1.4 Tujuan	4
1.5 Manfaat	4
1.6 Sistematika Penulisan	5
2 LANDASAN TEORI	6
2.1 <i>Preprocessing Text</i>	6
2.2 Algoritma Nazief dan Andriani	6
2.3 <i>Term Frequency–Inverse Document Frequency TF-IDF</i>	7
2.4 <i>Youtube</i>	9
2.5 Algoritma <i>Naive Bayes Classifier</i>	9
2.6 Bahasa Pemrograman <i>Python</i>	10



2.7	Penelitian Terdahulu	10
3	METODOLOGI PENELITIAN	12
3.1	Tahap Perencanaan	12
3.2	Tahap Pengumpulan Data	13
3.3	Tahap Analisa dan Hasil	14
3.4	Tahap Dokumentasi	16
4	ANALISIS DAN HASIL	17
4.1	Pengumpulan Data	17
4.2	<i>Cleaning</i>	17
4.3	Pelabelan	18
4.4	<i>Text Processing</i>	19
4.5	<i>Term Frequency-Inverse Document Frequency (TF-IDF)</i>	21
4.6	<i>Naive Bayes Classifier (NBC)</i>	22
4.6.1	Pembagian Data Latih dan Data Uji pada NBC	22
4.6.2	Hasil <i>K-Fold Cross Validation</i> pada NBC	22
4.6.3	Hasil Prediksi Kelas Data Menggunakan NBC	23
4.7	Analisis Klasifikasi	23
4.8	Visualisasi Data	24
4.9	Hasil Analisis Penelitian	26
5	PENUTUP	27
5.1	Kesimpulan	27
5.2	Saran	27
	DAFTAR PUSTAKA	
	LAMPIRAN A PELABELAN	A - 1
	LAMPIRAN B DAFTAR STOWORDS	B - 1
	LAMPIRAN C SURAT PERNYATAAN PAKAR	C - 1

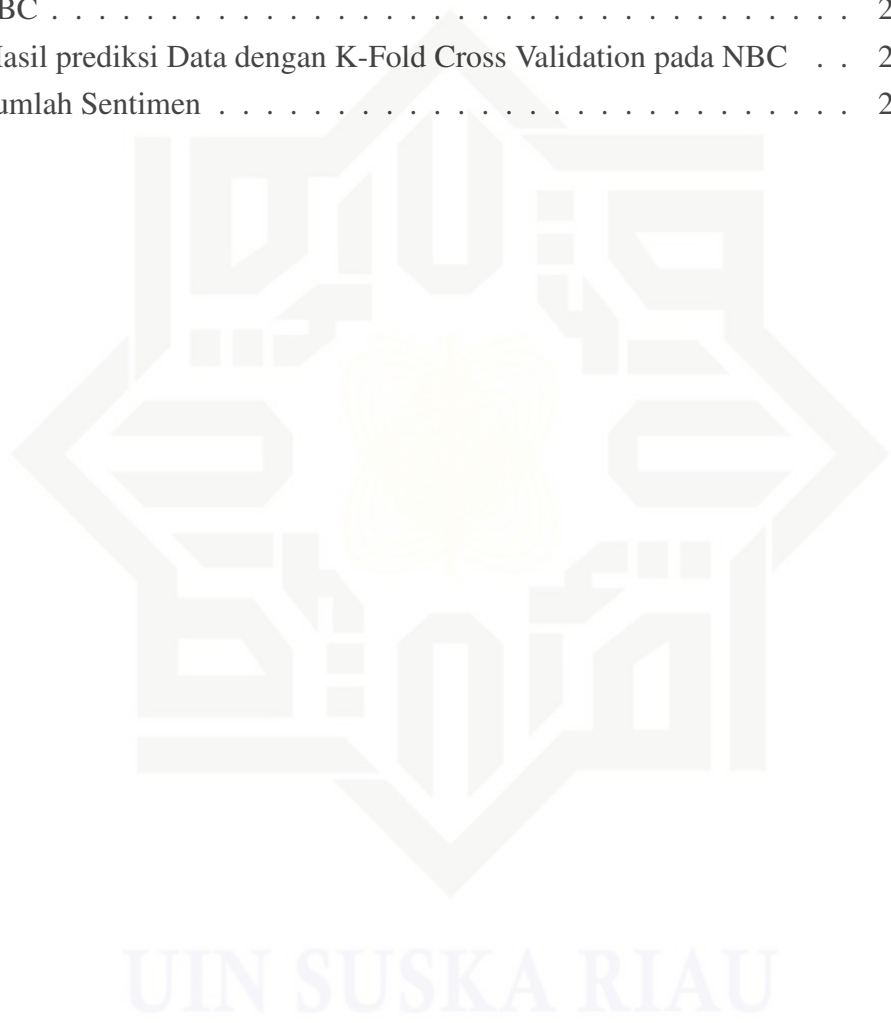
DAFTAR GAMBAR

2.1	Kerangka Penelitian Terdahulu	11
3.1	Alur Metodologi Penelitian	12
3.2	Alur <i>Preprocessing</i>	14
3.3	Alur <i>Naive Bayes Classifier</i>	15
4.1	Hasil Pembagian Data pada NBC dengan <i>K-Fold Cross Validation</i> .	22
4.2	<i>Wordcloud</i> hasil positif	24
4.3	<i>Wordcloud</i> hasil netral	25
4.4	<i>Wordcloud</i> sentimen negatif	25



DAFTAR TABEL

4.1	Data Awal	17
4.2	Data <i>Cleaning</i>	18
4.3	Pelabelan	19
4.4	Daftar Kata <i>Stopwords</i>	20
4.5	Hasil <i>Text Processing</i>	20
4.6	Hasil pembobotan TF-IDF	21
4.7	Hasil Pembagian Data Latih dengan <i>K-Fold Cross Validation</i> pada NBC	23
4.8	Hasil prediksi Data dengan K-Fold Cross Validation pada NBC	23
4.9	Jumlah Sentimen	24



DAFTAR SINGKATAN

BNPB	:	Badan Nasional Penanggulangan Bencana Alam
NBC	:	<i>Naive Bayes Classifier</i>
TF-IDF	:	<i>Term Frequency-Inverse Document Frequency</i>
BPBD	:	Badan Penanggulangan Bencana Daerah
IDF	:	<i>Inverse Dokumen Frequency</i>
DF	:	<i>Document Frequency</i>



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 1

PENDAHULUAN

1.1 Latar Belakang

Perkembangan teknologi yang berjalan dengan pesat memberikan pengaruh yang besar terhadap kehidupan masyarakat, salah satunya dalam hal penggunaan aplikasi sosial media. Penggunaan aplikasi sosial media di kalangan masyarakat sudah menjadi bagian penting dalam kehidupan sehari-hari dan seolah-olah interaksi telah dipindahkan ke dalam sebuah *platform virtual* (Alalwan, Rana, Dwivedi, dan Algharabat, 2017).

Pada saat ini media sosial telah menjadi alat komunikasi yang sangat populer di kalangan pengguna internet di Indonesia, media sosial merupakan salah satu tempat atau media untuk berekspresi dan berpendapat tentang berbagai macam topik salah satunya *Youtube*. Hartoko (2019) menyatakan bahwa “Saat ini siapa yang tak mengenal *Youtube* situs video ini makin “*booming*” di Indonesia, situs yang didirikan pada bulan Februari 2005 ini terus menanjak penggunaannya di seluruh dunia hingga membuat pemilik *Google* tergoda untuk membelinya pada tanggal 9 Oktober 2006”.

Interaksi di dalam sosial media dapat dilihat dari komentar-komentar yang tersedia di berbagai platform sebagai umpan balik dari setiap kegiatan yang ada di media sosial, mulai dari status yang berupa teks, gambar maupun video (Marleny, 2020). Komentar-komentar tersebut banyak terdapat opini yang terkadang positif maupun negatif, opini tersebut dapat dikaitkan dengan respon yang beragam, semakin banyak komentar maka semakin banyak pengguna yang berinteraksi, keadaan ini kadang menimbulkan interaksi yang memiliki respon tidak terduga.

Dari berbagai respon pada kolom komentar tersebut diperoleh sebuah informasi dari data yang tidak terstruktur sehingga perlu adanya suatu teknik untuk mendefinisikan nilai informasi sehingga dapat lebih terstruktur. Dari data yang telah diestrak didapatkan suatu informasi tentang opini atau pendapat dari pengguna media sosial terhadap entitas tertentu (Nugroho, 2017).

Menganalisis pemantauan opini publik di Internet atau media sosial untuk pengumpulan informasi opini public berupa analisis pencarian, pengelompokan informasi, analisis ucapan, dan analisis prediksi tren yang selanjutnya memverifikasi kebenaran dari nilai suatu informasi (Yasah, 2020). Wilayah Indonesia terletak di daerah *ring of fire* dan megatrast sehingga rentan bencana alam. Kejadian bencana alam terus secara grafis menunjuk *trend* yang terus meningkat.

Bencana alam merupakan peristiwa atau kejadian yang mengancam atau



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.

mengganggu kehidupan manusia yang disebabkan oleh alam, antara lain berupa gunung meletus, gempa bumi, tsunami, tanah longsor, banjir, kekeringan, dan angin topan (Fatimahsyam, 2018). Indonesia termasuk negara dengan tingkat bencana alam yang cukup tinggi, hal ini disebabkan oleh letak wilayah negara Indonesia yang berada di antara pertemuan tiga lempeng tektonik dunia, yaitu Eurasia, Indo Australia dan Pasifik (Ramadhan dan Prihandoko, 2017).

Berdasarkan catatan dari Badan Nasional Penanggulangan Bencana (BNPB), ada 2.571 bencana yang terjadi di Indonesia sepanjang tahun 2018. Bencana tersebut terdiri dari puting beliung sebanyak 804 kejadian, banjir 679 kejadian, tanah longsor 473 kejadian, kebakaran hutan dan lahan 370 kejadian, kekeringan 129 kejadian, letusan gunung api 58 kejadian, gelombang pasang dan abrasi 34 kejadian, gempabumi 23 kejadian, gempabumi dan tsunami 1 kejadian, dan tsunami 1 kejadian (BNPB, n.d.).

Peristiwa atau kejadian seperti bencana alam tersebut memicu masyarakat khususnya pengguna *Youtube* untuk memberikan informasi atau opini mereka dalam bentuk komentar. Komentar tersebut dapat berupa ungkapan rasa sedih, peduli, atau keluhan. Pengolahan data dari komentar *Youtube* ini akan menciptakan suatu trend yang dapat digunakan untuk kebutuhan informasi seperti pada bidang pendidikan, ekonomi, politik, dan lainnya. Kumpulan data kicauan tersebut dapat diolah dengan menggunakan konsep *text mining* (Reza, 2018)).

Text Mining merupakan bidang multi-disiplin yang mempelajari tentang pengambilan informasi, analisis teks, ekstraksi informasi, kategorisasi, pengelompokan, visualisasi, *data mining*, dan *machine learning* (Dang, 2014). *Text Mining* diperlukan untuk mengklasifikasikan komentar video *Youtube* berdasarkan topik penanggulangan bencana alam di Indonesia. Topik ini dipilih dengan tujuan agar penelitian ini dapat memberikan wawasan kepada masyarakat mengenai analisis sentimen masyarakat terhadap bencana alam di Indonesia dalam sosial media, khususnya *Youtube*.

Text Mining cenderung mengarah pada bidang penelitian *data mining*. Oleh karena itu, tidak mengherankan bahwa *text mining* dan *data mining* memiliki banyak kesamaan dalam bidang arsitektur (Feldman dan Sanger, 2007). Salah satu teknik *data mining* yang bisa digunakan dalam *text mining* adalah klasifikasi.

Metode *Naive Bayes Classifier* (NBC) menempuh dua tahap dalam proses klasifikasi teks, yaitu tahap pelatihan dan tahap klasifikasi. Pada tahap pelatihan dilakukan proses analisis terhadap sampel dokumen berupa pemilihan *vocabulary*, yaitu kata yang mungkin muncul dalam koleksi dokumen sampel yang sedapat mungkin dapat menjadi representasi dokumen. Selanjutnya adalah penentuan prob-



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.

abilitas prior bagi tiap kategori berdasarkan sampel dokumen. Pada tahap klasifikasi ditentukan nilai kategori dari suatu dokumen berdasarkan term yang muncul dalam dokumen yang diklasifikasi.

NBC adalah metode sederhana tetapi memiliki akurasi dan permomansi yang tinggi dalam pengklasifikasian teks (Routray, Swain dan Prava, 2013). *NBC* menghitung probabilitas kelas berdasarkan distribusi kata-kata yang ada dalam dokumen (Medhat, Hassan, Korashy, 2014). *Naive Bayes Classifier* memiliki beberapa keunggulan seperti sederhana, cepat dan akurasi yang tinggi. Banyak peneliti telah melakukan klasifikasi sentimen dengan menggunakan *Naive Bayes Classifier*. Namun klasifikasi ini memiliki keterbatasan utama yang tidak mungkin selalu memenuhi asumsi independensi antara atribut. Dan ini mempengaruhi tingkat akurasi klasifikasi (Dhande Patnaik, 2014).

Saat ini peneliti ingin melakukan percobaan penelitian topik Bencana Alam dengan media *Youtube* dengan Algoritma *NBC* sebagaimana peneliti sebelumnya yang menggunakan algoritma *NBC* yaitu Implementasi Sentimen Analisis Komentar *Channel* Video Pelayanan Pemerintah di *Youtube* Menggunakan Algoritma *Naive Bayes* oleh (Saputra, Subhi, dan Winatama, 2019). Hasil akurasi klasifikasi algoritma *Naive Bayes Classifier* pada *Channel* KemenPUPR menghasilkan nilai 69,23% persen dan pada *channel* Kemenkeu menghasilkan nilai 64,10%.

Selanjutnya penelitian tentang Analisis Sentimen Menggunakan *Naive Bayes Classifier* dengan *Chi-Square Feature Selection* Terhadap Penyedia Layanan Telekomunikasi oleh (Nisa, 2019). Hasil penelitian ini memberikan akurasi 85,5%, dan *f1-score* 84%, sedangkan tanpa menggunakan seleksi fitur *chi square* didapatkan akurasi 84,4%, dan *f1-score* 82%.

Kemudian penelitian tentang “Analisis Sentimen Tentang Opini Film pada Dokumer Twitter Berbahasa Indonesia Menggunakan *Naive Bayes* dengan Perbaikan Kata Tidak Baku” oleh (Antinasari et al., 2017). Penelitian ini menganalisa opini Twitter dan membaginya menjadi dua kelas yaitu positif dan negatif yang menghasilkan akurasi hingga 91.67%. Selanjutnya penelitian “*Scalable Sentiment Classification For Big Data Analysis Using Naive Bayes Classifier*” oleh (Liu et al., 2013). Hasil penelitian ini dapat berjalan dengan baik pada big data dan menghasilkan akurasi sebesar 82%.

Berdasarkan permasalahan di atas dan beberapa penelitian terdahulu, maka pada penelitian Tugas Akhir ini akan dilakukan penerapan *naive bayes classifier* untuk analisis sentimen tentang penanggulangan bencana alam di Indonesia.



1.2 Rumusan Masalah

Berdasarkan latar belakang di atas maka rumusan masalahnya adalah bagaimana menerapkan *Naive Bayes Classifier* untuk analisis sentimen masyarakat tentang penanggulangan bencana alam di Indonesia.

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 yang digunakan dalam penelitian ini merupakan data komentar video *Youtube* tentang Bencana Alam.
2. Algoritma yang digunakan untuk pengklasifikasian adalah *Naive Bayes Classifier*.
3. Data yang digunakan dengan rentang waktu Januari 2021 – Agustus 2021.
4. Tools yang digunakan untuk pengambilan dan klasifikasi data adalah bahasa pemrograman Python, dan Microsoft Excel.
5. Data sentimen dikumpulkan dari beberapa chanel *Youtube* yaitu CNN Indonesia, tvOnenews, metrotvnews, Top Videos of Uniqes and Strangers World, video on the spot terbaru.

1.4 Tujuan

Adapun tujuan dari penelitian ini adalah untuk:

1. Menguji akurasi algoritma *Naive Bayes Classifier* untuk analisis sentimen masyarakat tentang bencana alam di Indonesia dari data komentar video *Youtube* Bencana Alam.
2. Untuk menemukan informasi dan sentimen terkait komentar bencana alam

1.5 Manfaat

Adapun manfaat dari penelitian ini adalah:

1. Mendapatkan hasil akurasi algoritma *Naive Bayes Classifier* dari data komentar video *Youtube* tentang Bencana Alam.
2. Penelitian ini menambah wawasan kepada publik mengenai analisis sentimen masyarakat terhadap komentar bencana alam pada *Youtube*.
3. Mendapatkan visualisasi dari komentar sesuai dengan banyak kata guna analisa lebih lanjut terhadap kata yang sering muncul.
4. Mendapatkan informasi dari masyarakat sebagai analisa lebih lanjut untuk organisasi terkait seperti BPBD tentang Bencana Alam.

1.6 Sistematika Penulisan

Sistematika penulisan laporan Tugas Akhir ini disusun agar pembuatan laporan dapat lebih terstruktur dan lebih mudah dalam memahami penelitian yang dilakukan. Adapun sistematika penulisan tersebut adalah sebagai berikut:

BAB 1. PENDAHULUAN

BAB 1 pada tugas akhir ini berisi tentang: (1) Latar Belakang; (2) Perumusan Masalah; (3) Batasan Masalah; (4) Tujuan; (5) Manfaat; dan (6) Sistematika Penulisan. 4

BAB 2. LANDASAN TEORI

BAB 2 pada tugas akhir ini berisi tentang: (1) *Preprocessing Text*; (2) Algoritma Nazief dan Andriani; (3) *Term Frequency–Inverse Document Frequency (TF-IDF)*; (4) *Youtube*; (5) Algoritma *Naive Bayes*; (6) Bahasa Pemrograman *Python*; dan (7) Penelitian Terdahulu.

BAB 3. METODOLOGI PENELITIAN

BAB 3 pada tugas akhir ini berisi tentang: (1) Tahap Perencanaan; (2) Tahap Pengumpulan Data; (3) Tahap Hasil dan Analisis; dan (4) Tahap Dokumentasi.

BAB 4. ANALISA DAN HASIL

BAB 4 pada tugas akhir ini berisi tentang: (1) Pengumpulan Data; (2) *Cleaning Data*; (4) Pelabelan Data; (5) *Text Preprocessing*; (6) *Term Frequency–Inverse Document Frequency (TFIDF)*; (7) Algoritma *Naive Bayes*; (8) Analisis Klasifikasi; dan (9) Hasil Analisis Penelitian.

BAB 5. PENUTUP

pada tugas akhir ini berisi tentang: (1) Kesimpulan dan (2) Saran.

BAB 2

LANDASAN TEORI

2.1 *Preprocessing Text*

Menurut Feldman dan Sanger (2007) *text preprocessing* merupakan tahapan awal dari proses *text mining* terhadap teks yang bertujuan mengubah teks menjadi data yang dapat diolah selanjutnya. Sekumpulan karakter atau teks yang bersambung harus dipisah terlebih dahulu agar menjadi unsur yang lebih berarti. Suatu dokumen dapat dipisah menjadi bab, sub-bab, paragraf, kalimat, kata hingga suku kata.

Pada tahap *Text Preprocessing* ini akan dilakukan beberapa proses. Proses-proses yang dilakukan adalah sebagai berikut:

1. *Tokenizing*

Pada proses *tokenizing* sekumpulan karakter dalam suatu teks di pecah ke dalam satuan kata (Nurzahputra dan Muslim, 2016). Di dalam proses ini teks dipisah menjadi perkata berdasarkan spasi yang ditemukan.

2. *Filtering*

Menurut Nurzahputra dan Muslim (2016), proses *filtering* adalah proses menghilangkan kata-kata yang tidak mengandung makna atau sering disebut sebagai *stopword*. *Stopword* merupakan kosa kata yang tidak memiliki ciri dari suatu dokumen

3. *Stemming*

Proses *stemming* merupakan proses mengubah bentuk dari suatu kata menjadi bentuk kata dasarnya. Proses ini memiliki tujuan untuk mengilangkan imbuhan-imbuhan yang ada pada setiap kata baik berupa prefiks, sufiks, maupun konfiks (Nurzahputra dan Muslim, 2016).

2.2 *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 atau *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 (Wahyudi, Susyanto, dan Nugroho, 2017). Langkah-langkah pada algoritma *stemming* yaitu 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 3a: 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 3b. 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 diizinkan b. Awalan yang dideteksi sama dengan awalan yang dihilangkan sebelumnya. 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*.

2.3 Term Frequency–Inverse Document Frequency TF-IDF

Term Frequency-Inverse Document Frequency (TF-IDF) merupakan metode untuk memberikan bobot pada setiap kata untuk menentukan hubungan antar kata (*term*) terhadap dokumen (Herwijayanti, Ratnawati, dan Muflikhah, 2018). TF-IDF menggabungkan dua konsep dalam menghitung bobot sebuah kata, yaitu frekuensi kemunculan kata dalam sebuah dokumen tertentu dan inverse frekuensi dokumen yang mengandung kata tersebut.

Frekuensi kemunculan kata dalam dokumen menunjukkan seberapa penting kata tersebut dalam sebuah dokumen. Sedangkan, frekuensi dokumen yang mengandung kata tersebut menunjukkan bahwa kata tersebut adalah kata umum. Jadi, bobot hubungan antara sebuah dokumen dan sebuah kata akan memiliki bobot yang tinggi apabila frekuensi kata tersebut tinggi dalam sebuah dokumen dan frekuensi keseluruhan dokumen yang mengandung kata tersebut rendah (Wahyuni, Prastiyan-

to, dan dan Suprpto, 2017).

Tahap-tahap dalam melakukan pembobotan kata adalah sebagai berikut:

1. *Term Frequency* (TF)

Term Frequency merupakan frekuensi kemunculan kata pada suatu dokumen teks. *Term Frequency* $t_{ft,d}$ didefinisikan jumlah kemunculan *term* t pada dokumen d . Persamaan dari *Term Frequency* $t_{ft,d}$ dapat dilihat pada Persamaan 2.1.

$$W_{t.ftd} = \begin{cases} 1 + \log_{10} t_{ft,d} & t_{ft,d} > 0 \\ 0 & t_{ft,d} = 0 \end{cases} \quad (2.1)$$

Keterangan:

$t_{ft,d}$ = Jumlah kemunculan term t pada dokumen d .

2. *Document Frequency* (DF)

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

3. *Invers Document Frequency* (IDF)

Invers Document Frequency merupakan frekuensi kemunculan *term* pada keseluruhan dokumen teks. *Term* yang jarang muncul pada keseluruhan dokumen teks memiliki nilai *Invers Document Frequency* lebih besar dibandingkan dengan term yang sering muncul. Persamaan dari *Invers Document Frequency* (IDF) dapat dilihat pada Persamaan 2.2.

$$id_{ft} = \log_{ft} \frac{N}{df_t} \quad (2.2)$$

Keterangan:

N = Jumlah dokumen teks.

id_{ft} = 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 dapat dilihat pada Persamaan 2.3.

$$W_{t,d} = W_{t.ftd} \times id_{ft} \quad (2.3)$$

Keterangan:

$W_{t.ftd}$ = Term Frequency.

idf_t = Invers Dokument Frequency.

2.4 Youtube

YouTube merupakan media sosial video online terbesar di dunia yang digunakan untuk menampilkan berbagai macam video yang dibuat oleh pengguna dan perusahaan di bidang konten media (Clement, n.d.).

Youtube merupakan salah satu layanan dari *Google* yang memfasilitasi penggunaannya untuk mengupload dan mengunduh video dan bisa diakses oleh pengguna lain secara gratis. “Bisa dikatakan *Youtube* adalah *database* video yang paling populer di dunia internet, bahkan yang paling lengkap dan variatif (Nanuru, *Youtube*, and Modern, n.d.)”. “Aplikasi *Youtube* sudah ada pada Februari 2005, didirikan oleh tiga orang mantan karyawan *Paypal* yaitu Steven Chen, Jawed Karim dan Chad Hurley Yang berkantor pusat di San Bruno, California (Diurna and Vi, 2017)”.

Kehadiran *Youtube* sangat membawa pengaruh yang luar biasa kepada masyarakat, khususnya yang gemar membuat video blog (*vlog*), film pendek dan karya lainnya. *Youtube* bisa digunakan untuk mempublikasikan karya yang telah dibuat tanpa memerlukan biaya yang tinggi, hanya bermodalkan *gadget* yang kompatibel. *Content creator* merupakan sebutan untuk orang yang membuat konten berupa video dan mengunggahnya pada *YouTube*. Di *Youtube* juga disediakan fitur komentar sebagai penyalur respon masyarakat terhadap konten video yang di tontonnya.

Berikut *link* video *youtube* yang dijadikan untuk pengambilan *dataset*:

1. https://youtu.be/0Xf6BcFwl_Q (*Top Videos of Uniques and Strangers World*)
2. https://youtu.be/y_QBPw_iVME (CNN Indonesia)
3. <https://youtu.be/u4gdziFtoOU> (*Top Videos of Uniques and Strangers World*)
4. <https://youtu.be/KXHq-GwokQg> (tvOnenews)
5. <https://youtu.be/NYcbwdb1H68> (CNN Indonesia)
6. <https://youtu.be/b6-WHjZDLDU> (tvOnenws)
7. <https://youtu.be/TRWgvBipuHQ>
8. <https://youtu.be/ivX1S6cFUCs> (video on the spot terbaru)
9. <https://youtu.be/ko38GDspVJE> (tvOnenews)
10. https://youtu.be/tt_rD2dhQVg (metrotvnews)

2.5 Algoritma Naive Bayes Classifier

Algoritma *Naive Bayes* merupakan salah satu algoritma yang termasuk kedalam klasifikasi. Metode ini adalah metode terbaik dalam klasifikasi yang diter-

mukan oleh Thomas Bayes dengan memprediksi peluang dimasa depan berdasarkan data dimasa sebelumnya (Sari and Hayuningtyas, 2019). *Naive Bayes Classifier* dikembangkan oleh Reverend Thomas Bayes pada abad ke 18. *Naive Bayes Classifier* menerapkan fungsi statistik sederhana berdasarkan teorema Bayes dengan asumsi keberadaan dari suatu fitur tertentu terhadap suatu kelas yang tidak berhubungan dengan fitur lainnya.

Naive Bayes Classifier merupakan suatu metode yang menggunakan perhitungan probabilitas (Sari and Hayuningtyas, 2019). Metode *Laplace Smoothing* adalah metode untuk menghindari hasil klasifikasi bernilai 0 dikarenakan data *Testing* tidak ditemukan pada data *training*, dengan menggunakan metode *Laplace Smoothing* yang paling sederhana dengan hanya menambahkan angka 1 (satu) metode *Laplace Smoothing* memiliki *Performance* yang cukup baik (Listiwarni, 2019).

2.6 Bahasa Pemrograman Python

Python adalah bahasa pemrograman yang berorientasi objek yang biasanya digunakan untuk keperluan pengembangan perangkat lunak, serta dapat berjalan di berbagai *Operating System*. *Python* merupakan bahasa pemrograman yang tidak ada batasan dalam penyalinan atau pendistribusiannya. Beberapa fitur yang dimiliki *Python* adalah untuk (Perkasa, Widyantara, dan Susanto, 2014):

1. Memiliki *library* yang luas.
2. Memiliki bahasa yang *user friendly*.
3. Memiliki layout yang mudah dalam hal pengecekan, pembacaan kembali dan penulisan ulang *sourcecode*,
4. Berorientasi objek.

2.7 Penelitian Terdahulu

Implementasi Sentimen Analisis Komentar *Channel* Video Pelayanan Pemerintah di *Youtube* Menggunakan Algoritma *Naive Bayes* oleh (Saputra dkk., 2019). Hasil akurasi klasifikasi algoritma *Naive Bayes Classifier* pada Channel KemenPUPR menghasilkan nilai 69,23% persen dan pada *channel* Kemenkeu menghasilkan nilai 64,10%.

Selanjutnya penelitian tentang Analisis Sentimen Menggunakan *Naive Bayes Classifier* dengan *Chi-Square Feature Selection* Terhadap Penyedia Layanan Telekomunikasi oleh (Nisa, 2019). Hasil penelitian ini memberikan akurasi 85,5%, dan *f1-score* 84%, sedangkan tanpa menggunakan seleksi fitur *Chi Square* didapatkan akurasi 84,4%, dan *f1-score* 82%.

Kemudian penelitian tentang “Analisis Sentimen Tentang Opini Film pa-

da Dokumer Twitter Berbahasa Indonesia Menggunakan *Naive Bayes* dengan Perbaikan Kata Tidak Baku” oleh (Antinasari, Perdana, dan Fauzi, 2017). Penelitian ini menganalisa opini twitter dan membaginya menjadi dua kelas yaitu positif dan negatif yang menghasilkan akurasi hingga 91.67%.

Gambar 2.1.



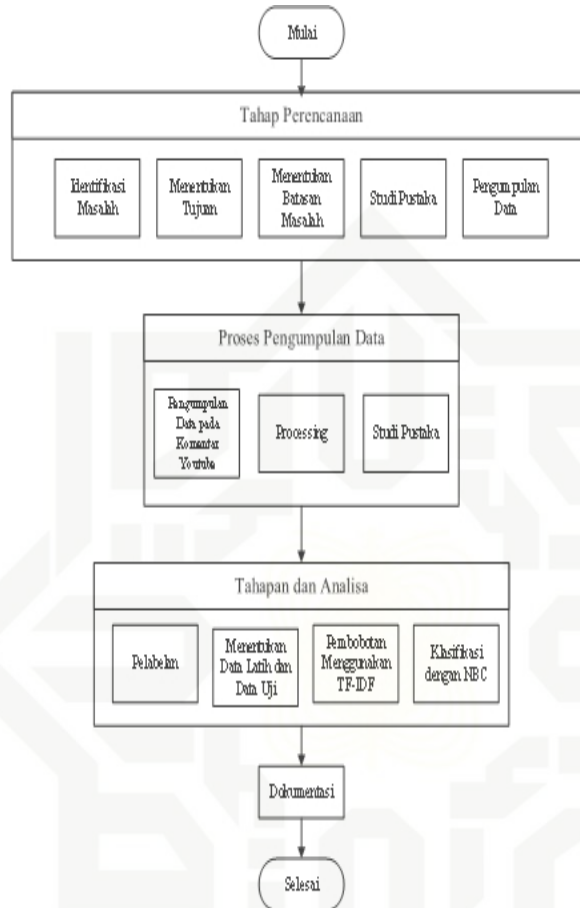
Gambar 2.1. Kerangka Penelitian Terdahulu

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 ini merupakan langkah-langkah dalam penelitian Tugas Akhir dapat dilihat pada Gambar 3.1.



Gambar 3.1. Alur Metodologi Penelitian

Berikut ini adalah penjelasan langkah-langkah metodologi penelitian tugas akhir:

3.1 Tahap Perencanaan

Tahap perencanaan merupakan tahapan-tahapan yang harus ada saat akan melakukan penelitian, tahapan yang direncanakan pada penelitian ini adalah :

1. Identifikasi Masalah

Mengidentifikasi permasalahan yang terjadi pada media sosial *Youtube*. Penentuan masalah memiliki tujuan untuk mengamati permasalahan yang terjadi tentang bencana alam berdasarkan opini masyarakat yang berupa komentar *Youtube*.

2. Menentukan Tujuan
Menentukan tujuan berfungsi untuk menjelaskan tujuan dan kerangka dari penelitian yang akan dilaksanakan. Tujuan dari penelitian ini adalah melakukan klasifikasi opini masyarakat pada data yang didapat dari ulasan *Youtube*. Lalu menghitung akurasi algoritma yang diterapkan pada data yaitu *Naive Bayes Classifier* (NBC).
3. Menentukan Batasan Masalah
Menentukan batasan masalah bertujuan agar tetap berfokus pada objek penelitian sehingga tidak keluar dari cakupan objek penelitian yang telah ditentukan.
4. Studi Pustaka
Studi pustaka bertujuan untuk mengetahui teori-teori yang digunakan untuk menyelesaikan permasalahan dalam penelitian. Studi pustaka juga berfungsi untuk mendapatkan dasar-dasar rujukan yang kuat bagi penelitian untuk menganalisis opini masyarakat tentang media massa pada *Youtube* menggunakan algoritma NBC.

3.2 Tahap Pengumpulan Data

Pengumpulan data yang dilakukan melalui beberapa tahapan sebagai berikut:

1. Pengumpulan Data *Youtube* (*Scrapping*)
Pada tahapan pengumpulan data, peneliti mengambil data melalui *Youtube* yaitu berupa data komentar yang mengarah pada video bencana alam.
2. *Preprocessing*
Berikut adalah tahapan-tahapan dari *Preprocessing*:
 - (a) *Tokenizing*
Tahapan *Tokenizing*, yaitu tahap pemisahan kata berdasarkan karakter spasi pada setiap dokumen.
 - (b) *Filtering*
Tahapan *Filtering*, yaitu tahap membuang kata-kata yang tidak diperlukan dan proses *tokenizing*. Penghapusan kata dilakukan dengan menggunakan stopword atau kamus kata.
 - (c) *Stemming*
Proses *stemming* merupakan proses mengubah bentuk dari suatu kata menjadi bentuk kata dasarnya.

Alur dari proses *preprocessing* terhadap dokumen (*Youtube*) dapat dilihat pada Gambar 3.2.

3. Studi Pustaka Tambahan

Studi pustaka merupakan dasar teori dalam melakukan sebuah penelitian sehingga analisa yang dilakukan memiliki referensi yang kuat dalam menyelesaikan masalah.



Gambar 3.2. Alur *Preprocessing*

3.3 Tahap Analisa dan Hasil

Tahap analisa dan hasil memiliki beberapa tahapan 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 jelas.

2. Pembobotan

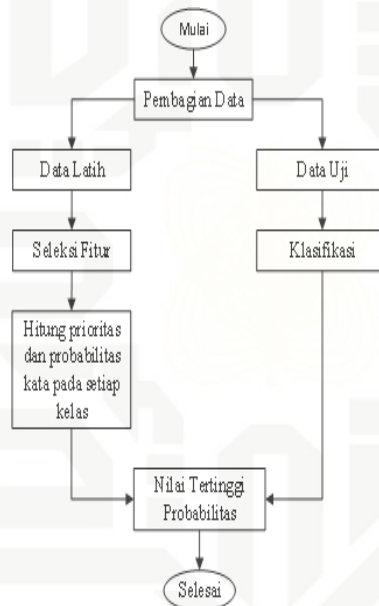
Tahap pembobotan adalah setiap kata diberi bobot dengan menggunakan *Term Frequency dan Inverse Dokument Frequency (TF-IDF)*.

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 *Naive Bayes Classifier (NBC)* setelah data tersebut diberi pembobotan.

4. Analisis Klasifikasi Naive Bayes Classifier (NBC)

Pada tahap ini dilakukan kalsifikasi menggunakan algoritma *Naive Bayes Classifier (NBC)*, kemudian dilanjutkan dengan menghitung akurasi algoritma dan visualisasi data. Pada klasifikasi *Naive Bayes Classifier (NBC)* terdapat beberapa langkah seperti Gambar 3.3.



Gambar 3.3. Alur *Naive Bayes Classifier*

5. Analisa Hasil Klasifikasi

Pada tahapan ini di hitung berapa hasil akurasi yang didapatkan untuk melihat berapa tingkat akurasi yang didapat menggunakan algoritma NBC. Kemudian memvisualisasikan kata dari dataset berdasarkan frekuensi kata yang muncul sesuai dengan kelas masing-masing yaitu netral, positif, dan negatif.

3.4 Tahap Dokumentasi

Tahap dokumentasi adalah tahap mendokumentasikan hasil dari penelitian yang telah dilaksanakan dengan melakukan pembuatan laporan. Hasil akhir pada tahapan ini berupa dokumentasi Laporan Tugas Akhir.

Hak cipta milik UIN Suska Riau

State Islamic University of Sultan Syarif Kasim 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. Penerapan algoritma Naive Bayes Classifier menghasilkan akurasi sebesar 55,71%
2. Sentimen terkait data komentar youtube bencana alam pada kelas tertinggi adalah kelas negatif sebanyak 585. Pada kelas netral sebanyak 167 dan kelas positif sebanyak 8.

5.2 Saran

Penelitian ini tentunya memiliki kekurangan, khususnya terkait dengan processing text seperti mengubah singkatan ke bentuk dasar agar data teks yang diperoleh menjadi lebih baku lagi. Selanjutnya, melakukan validasi label positif, negatif, dan netral secara manual agar mendapatkan hasil yang optimal, serta melakukan percobaan algoritma yang berbeda dan menggunakan metode klasifikasi lainnya.

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 PUSTAKA

- Alalwan, A. A., Rana, N. P., Dwivedi, Y. K., dan Algharabat, R. (2017). Social media in marketing: A review and analysis of the existing literature. *Telematics and Informatics*, 34(7), 1177–1190.
- Antinasari, P., Perdana, R. S., dan Fauzi, M. A. (2017). Analisis sentimen tentang opini film pada dokumen twitter berbahasa indonesia menggunakan naive bayes dengan perbaikan kata tidak baku. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 1(1), 1–28.
- BNPB, A. (n.d.). *Jumlah bencana alam, 2018*. Retrieved from <http://bnpb.cloud/dibi/laporan5>
- Clement, J. (n.d.). *Statista, 2021*. Retrieved from <https://www.statista.com/topics/2019/youtube//>
- Dang, S. (2014). Text mining : Techniques and its application. *International Journal of Engineering and Technology innovation*, 12(1), 22–25.
- Fatimahsyam, F. (2018). Pengintegrasian pengurangan risiko bencana dengan pendekatan mazhab antroposentris. *SUBSTANTIA*, 20(1), 49–65.
- Feldman, R., dan Sanger, J. (2007). *The text mining handbook: advanced approaches in analyzing unstructured*. Cambridge University Press.
- Hartoko, A. (2019). *Rahasia cepat tenar dan dapat duit lewat youtube*. PT Elex Media Komputindo.
- Herwijayanti, B., Ratnawati, D. E., dan Muflikhah, L. (2018). Klasifikasi berita online dengan menggunakan pembobotan tf-idf dan cosine similarity. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 2(1), 306–312.
- Marleny, F. (2020). Analisis pengguna whatsapp terhadap kesalahan mengirim pesan teks menggunakan metode klasifikasi. *Jurnal Teknologi Informasi Universitas Lambung Mangkurat (JTIULM)*, 5(1), 19–24.
- Nisa. (2019). Analisis sentimen menggunakan naïve bayes classifier dengan chi-square feature selection terhadap penyedia layanan telekomunikasi. Dalam *e-proceeding of engineering* (hal. 104–108).
- Nugroho, H. (2017). Pengaruh algoritma stemming nazief-adriani terhadap kinerja algoritma winnowing untuk mendeteksi plagiarisme bahasa indonesia. *Jurnal ULTIMA Computing*, 9(1), 11–38.
- Nurzahputra, A., dan Muslim, M. (2016). Analisis sentimen pada opini mahasiswa menggunakan natural language processing. Dalam *Seminar nasional ilmu komputer* (hal. 114–118).
- Ramadhan, dan Prihandoko. (2017). Penerapan data mining untuk analisis data



bencana milik bnpb menggunakan algoritma k-means dan linear regression. *Jurnal Information dan Komputer*, 22(1), 50–65.

Reza, N. G. I. (2018). *Penerapan algoritma dbscan untuk pencarian trend topik pilkada pekanbaru 2017 pada twitter* (Unpublished doctoral dissertation). Skripsi. UIN Sultan Syarif Kasim.

Saputra, P. Y., Subhi, D. H., dan Winatama, F. Z. A. (2019). Implementasi analisis komentar channel video pelayanan pemerintah di youtube menggunakan algoritma naïve bayes. *Jurnal Informatika Polinema*, 1(1), 1–28.

Tala, F. (2003). *A study of stemming effects on information* (Unpublished doctoral dissertation). Skripsi. Institute for Logic, Language and Computation Universiteit van Amsterdam The Netherlands.

Wahyudi, D., Susyanto, T., dan Nugroho, D. (2017). Implementasi dan analisis algoritma stemming nazief dan adriani dan porter pada dokumen berbahasa indonesia. *Jurnal Ilmiah SINUS*, 15(2), 49–56.

Wahyuni, R. T., Prastiyanto, D., dan dan Suprpto, E. (2017). Penerapan algoritma cosine similarity dan pembobotan tf-idf pada sistem klasifikasi dokumen skripsi. *Jurnal Teknik Elektro*, 9(1), 18–23.

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 A

PELABELAN

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.

No	Komentar	Label Pakar
1	udahannya bakal subur banget tuh pasti soalnya tanah vulkanis	positif
2	weewwww	netral
3	akainu	netral
4	gini sih jdinya klw tidak ada jalur pembuangannya seperti di indonesia	netral
5	ada jalur larvanya	netral
6	ya allah aku gak mau ma i	netral
7	ya allah sungguh besar kekuasaan mu ya allah aku sangat takjup dengan	positif
8	kekuasaan mu yang sangat besar itu masyaallah	netral
9	astagfirulloh	netral
10	semoga kita selalu di lindungi allah swt amin	netral
11	astagfirullah	netral
12	gw kirain di	netral
13	apakah tanda tanda kiamat na uji bilah	positif
14	ya allah alhamdulillah gk di indonesia	netral
15	kebanyakan iklan	netral
16	masyallah gimna neraka ya allah	negatif
17	gw kepikiran hewan disana terkhusus kucing	netral
18	kalo sudah begini apa yg di banggakan coba	netral
19	itulah mengapa aku trauma nonton berita semoga kita di beri kesehatan	netral
20	dan rezeki dilancarkan amin	netral
21	awal mulah kiamat	negatif
22	wih seru tuh ekwkekkw canda	netral
23	ngeri banget ya gaess	negatif
24	hater bilang ini editan	netral
25	fiks ulah akainu	netral
26	ada admiral akainu	netral
27	semoga kameramen dalam lindungan yang maha kuasa	positif
28	ngeri hampir sama kayak scene hawai film	negatif
29	valir	netral
30	seperti bayangan neraka ya allah	netral
31	anjay kameramennya santai bet	netral
32	meng gg	netral
33	takut kayak pilem ini bukan pilem gt lt	negatif
34	keramennya deket bamget	netral
35	azab	netral
36	maap nih salfok ama mbaknya yang baca r nya kek medok gimana gitu	netral
37	berita lengkapnya terbaru a a	positif
38	yg merasa hebat	positif

Table A.1

No	Komentar	Label Pakar
36	b i the real american movies i b	netral
37	itu baru bencana alam kalo nanti kiamat	negatif
38	keren ya gunung berapi di luar negeri meletus mengeluarkan lahar aja pelan sampe bisa di lihat gitu coba kalo di indonesia meletus dikit aja langsung memuntahkan batubatuan dan awan panas abu vulkanik beda jauh ama luar negeri cuma mengeluarkan lahar dan mengalir pelan pelan tanpa abu vulkanik gempa bahkan wedus gembel awan panas	negatif
39	itu tapi setelah itu biasanya tanah nya jadi subur	positif
40	apakah indonesia sudah siap kalo menghadapi bencana seperti ini mrrerka masih bisa lari lah kita wedus gembel yg ngejar	netral
41	kirain di jogja tadi bisa aja bikin judul	netral
42	lari hah apa itu	netral
43	marijan mana marijan	netral
44	dunia sudah tua kita perbaiki diri perkemaskan diri kita selawat insyaallah mendapat syafaat baginda rasullullah	netral
45	menarik	positif
46	presenter nya mcam org malaysia	netral
47	masagi	netral
48	mungkin tuhan mulai bosan melihat tingkah manusia yg selalu salah dan bangga dengan dosa kita tanya pada rumput yang bergoyang	negatif
49	ini akan menjadi objek wisata aamin	netral
50	masya allah	netral
51	ya allah selamatkanlah warga di sana	netral
52	materi hanya lah titipan	netral
53	kekurangan berita sekali	negatif
54	solusinya sih bikin hujan buatan cepet reda	positif
55	keren	positif
56	kok bapak bapak nya santuy aja yak	netral
57	lava nya kok gak di stop pakai drone	netral
58	semoga semua manusia di sana mati	negatif
59	kampungnya pedri	netral
60	kayak di film	netral
61	ga penting berita luar negri masalah di indonesia lebi banyak	positif
62	ekspresi si kucing ketika di rekam ini videonya a a	netral
63	semoga kita semua umat di dunia selalu dlm lindungan tuhan yme amin	netral
64	sampai turun ke jalan	netral
65	ckenarrie ngakak w dengernyh hahh a a	netral
66	ngeri banget cuk	negatif
67	astagfirullah	netral
68	pertama kali muntahan lava di jalanan	negatif
69	resiko tinggal didekat gunung berapi	positif
70	semoga semuanya di berikan kesehatan oleh alloh aminn	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

No	Komentar	Label Pakar
71	nyata ada ny dngn plm volcano	netral
72	tempatny burung kenari	netral
73	ya allah ngeri liatnya	negatif
74	innalilahi wa innailaihi rojiuun	netral
75	mbayangno neroko ngeri	negatif
76	ini dmn min	netral
77	ya alloh semoga selalu di beri kesabaran pada warga semuanya ciri ciri kiamat besar adalah hutan mulai habis bukit banyak yang bolong seperti di bolongin untuk terowongan jalan banyak di arab sudah tergambar bukit pada di bolongin laut seringkali meluapkan emosinya tiba tiba gunung meluapkan amarahnya tergambar jelas alam ini sudah sangat marah kepada manusia zaman sekarang mereka	positif
78	menganggap manusia kini tidak pernah memperhatikannya selalu tamak sejatinya kiamat besar adalah akibat dari perbuatan manusia itu sendiri semua sudah terhitung seperti umur manusia cepat atau lambat akan lenyap kehidupan seperti film ada opening ada ending kita tengah di episode akhir hanya perlu waktu untuk menantikan endingnya	positif
79	ya allah	netral
80	semoga hewan juga bisa menyelamatkan diri dan tidak ada yg tersakiti ya allah	negatif
81	waw ngeri sekali sahabat	negatif
82	ini gunung yang di kepulauan canaria di baratnya maroko kan bukan di mainlandnya spanyol	netral
83	mantap masak gak perlu pake gas jika semua gunung diperintahkan allah mengeluarkan isi perutnya kemana lagi manusia yang sombong dan serakah akan mencari perlindungan	negatif
84	ingatkan hati selalu ingat dan taat perintah allah hanya karena rahmat dan kasih sayang allah kita masih baik baik saja ini hanya peringatan kecil dari allah supaya manusia selalu ingat bersyukur dan taat perintahnya	positif
85	ya allah hurobi sereem bgt	netral
86	semoga diberi kesabaran yang terkena musibah semua ini ujian dari tuhan	netral
87	akainu s power	netral
88	wedus gembell	netral
89	mirip kisah pompeii	netral
90	bukti keserakahan manusia	negatif
91	ya allah	netral
92	tak sangka clickbait tp liat chanelnya eh metritv	netral
93	kenapa g berapi indo sama luar negri beda lava diindo ga pernah datang ke rumah warga	positif
94	host nya lebay yah klk ngomong chanarii	netral
95	koment debat berulah	netral



Table A.1

No	Komentar	Label Pakar
96	gak heran klo rumah di luar negeri banyak yg dekat gunung biasa gunung berapi	positif
97	besok tinggal krakatau	netral
98	penyiarinya suaranya sok jd orang luar jd gk fokus	negatif
99	ya allah gusti	netral
100	masih ada yang ingat sama film ini dante s peak	netral
...		
7595	ayo bebenah pasang pohon penyerapan kepala yg bermanfaat buahnya untuk produk pabrik santan dan wisata pohon mangrove di perbanyak atas ijin allah swt memperbaiki hidup kita di alam dunia semangat pemuda indonesia allahuakbar	positif
7596	dampak manusia serakah hanya demi duniawi saya lahir di muara gembong desa pantai sederhana kec muara gembong saya tau persis kejadian soal terjadinya abrasi muara gembong dulu sebelum datang limbah dan pengerukan pasir hidup warga setempat makmur pasir bwt uruk ancol ngambil pasirnya dari muara gembong kalau	negatif
7597	di tangkap ma warga di bawa kapolsek terlepas tanpa proses karna gk sanggup dengan beking nya terus jaman ahok nguruk proyeknya pasir nya dari muara gembong juga blm limbah pabrik yg bwt warga semakin sulit mencari nafkah itu lah alam murka karna ulah manusia rakus salam anak desa	negatif
7598	dampak ulah manusia	netral
7599	subanaaallah semoga allah selalu memberikan kebahagiaan pada ibu suaamah dan lebih baik amin	positif
7600	ibarat laut itu air dalam piring apabila di tuangkan nasi air meluap kan makanya pesisir laut jgn di urug sehingga air laut meluap blm lagi nnt musim hujan tampah penuh air laut	netral
7601	ntar pindah aja ke pulau reklamasi ciptaan siapa gak taulah	netral
7602	wow	netral
7603	harusnya dulu jng dibuat tambak tapi ditanam mangrop yg banyak biar ga abrasi	positif
7604	komen kedua	netral

LAMPIRAN B

DAFTAR *STOWORDS*

No	Stopwords	No	Stopwords	No	Stopwords
1	ada	31	asalkan	61	begitu
2	adalah	32	atas	62	begitukah
3	adanya	33	atau	63	begitulah
4	adapun	34	ataukah	64	begitupun
5	agak	35	ataupun	65	bekerja
6	agaknya	36	awal	66	belakang
7	agar	37	awalnya	67	belakangan
8	aja	38	bagai	68	belum
9	akan	39	bagaikan	69	belumlah
10	akankah	40	bagaimana	70	benar
11	akhir	41	bagaimanakah	71	benarkah
12	akhiri	42	bagaimanapun	72	benarlah
13	akhirnya	43	bagi	73	berada
14	aku	44	bagian	74	berakhir
15	akulah	45	bahkan	75	berakhirilah
16	amat	46	bahwa	76	berakhirnya
17	amatlah	47	bahwasanya	77	berapa
18	anda	48	baik	78	berapakah
19	andalah	49	bakal	79	berapalah
20	antar	50	bakalan	80	berapapun
21	antara	51	balik	81	berarti
22	antaranya	52	banyak	82	berawal
23	apa	53	bapak	83	berbagai
24	apaan	54	baru	84	berdatangan
25	apabila	55	bawah	85	beri
26	apakah	56	beberapa	86	berikan
27	apalagi	57	begini	87	berikut
28	apatah	58	beginian	88	berikutnya
29	artinya	58	beginikah	89	berjumlah
30	asal	60	beginilah	90	berkali-kali
91	berkata	124	bolehkah	157	Diantara
92	berkehendak	125	bolehlah	158	Diantaranya
93	berkeinginan	126	buat	159	diberi
94	berkenaan	127	bukan	160	diberikan
95	berlainan	128	bukankah	161	diberikannya
96	berlalu	129	bukanlah	162	dibuat
97	berlangsung	130	bukannya	163	dibuatnya
98	berlebihan	131	bulan	164	didapat

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.

Table B.1

No	Stopwords	No	Stopwords	No	Stopwords
99	bermacam	132	bung	165	didatangkan
100	bermacam-macam	133	cara	166	digunakan
101	bermaksud	134	caranya	167	diibaratkan
102	bermula	135	cukup	168	diibaratkannya
103	bersama	136	cukupkah	169	diingat
104	bersama-sama	137	cukuplah	170	diingatkan
105	bersiap	138	cuma	171	diinginkan
106	bersiap-siap	139	dahulu	172	dijawab
107	bertanya	140	dalam	173	dijelaskan
108	bertanya-tanya	141	dan	174	dijelaskannya
109	berturut	142	dapat	175	dikarenakan
110	berturut-turut	143	dari	176	dikatakan
111	bertutur	144	daripada	177	dikatakannya
112	berujar	145	datang	178	dikerjakan
113	berupa	146	dekat	179	diketahui
114	besar	147	demi	180	diketahuinya
115	betul	148	demikian	181	dikira
116	betulkah	149	demikianlah	182	dilakukan
117	biasa	150	dengan	183	dilalui
118	biasanya	151	depan	184	dilihat
119	bila	152	di	185	dimaksud
120	bilakah	153	dia	186	dimaksudkan
121	bisa	154	diakhiri	187	dimaksudkannya
122	bisakah	155	diakhirinya	188	dimaksudnya
123	boleh	156	dialah	189	diminta
190	dimintai	223	ditunjuk	255	hingga
191	dimisalkan	224	ditunjuki	256	ia
192	dimulai	225	ditunjukkan	257	ialah
193	dimulailah	226	ditunjukkannya	258	ibarat
194	dimulainya	227	ditunjuknya	259	ibaratkan
195	dimungkinkan	228	dituturkan	260	ibaratnya
196	dini	229	dituturkannya	261	ibu
197	dipastikan	230	diucapkan	262	ikut
198	diperbuat	231	diucapkannya	263	ingat
199	diperbuatnya	232	diungkapkan	264	ingat-ingat
200	dipergunakan	233	dong	265	ingin
201	diperkirakan	234	dua	266	inginkah
202	diperlihatkan	235	dulu	267	inginkan
203	diperlukan	236	empat	258	ini
204	diperlukannya	237	enggak	259	inikah
205	dipersoalkan	238	enggaknya	270	inilah
206	dipertanyakan	239	entah	271	innalillahi

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 B.1

No	Stopwords	No	Stopwords	No	Stopwords
207	dipunyai	240	entahlah	272	itu
208	diri	241	ga	273	itukah
209	dirinya	242	gak	274	itulah
210	disampaikan	243	guna	275	jadi
211	disebut	243	gunakan	276	jadilah
212	disebutkan	244	hal	277	jadinya
213	disebutkannya	245	hampir	278	jangan
214	disini	246	hanya	279	janganlah
215	disinilah	247	hanyalah	280	janganlah
216	ditambahkan	248	hari	281	jauh
217	ditandakan	249	harus	282	jawab
218	ditanya	250	haruslah	283	jawaban
219	ditanyai	251	harusnya	284	jawabnya
220	ditanyakan	252	hendak	285	jelas
221	ditegaskan	253	hendaklah
222	ditujukan	254	hendaknya	782	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.

LAMPIRAN C

SURAT PERNYATAAN PAKAR

SURAT PERNYATAAN

Saya yang bertanda tangan dibawah ini:

Nama : Nuratika, S.Hum., M.Pd
Tempat/Tanggal Lahir : Sebalang/ 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 : Nurhalimah
NIM : 11753200031
Jurusan : Sistem Informatika
Judul Tugas Akhir : Penerapan *Naive Bayes Classifier* Untuk Analisis Sentimen
Tentang Penanggulangan Bencana Alam di Indonesia.

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.



© Hak cipta milik UIN SUSKA RIAU
State Islamic University of Sultan Syarif Kasim Riau

DAFTAR RIWAYAT HIDUP

Nurhalimah lahir di Simpang Tonang, JR. Tonang Raya, Kecamatan Dua Koto, Kabupaten Pasaman, Sumatera Barat, pada tanggal 14 Februari 1999, anak ke empat dari empat bersaudara dari Bapak Sahril dan Ibu Ratna Dewita. Penulis ini beralamatkan di jalan Asta Karya Blok E No 5, Kelurahan Tuah Madani, Kecamatan Tampan, Pekanbaru. Pengalaman pendidikan yang dilalui pada pendidikan Sekolah Dasar (SD) Negeri 05 Tonang Raya pada tahun 2006-2012, Sekolah Menengah Pertama (SMP) Negeri 1 Dua Koto pada tahun 2012-2014, dan Sekolah Menengah Atas (SMA) Negeri 1 Dua Koto pada tahun 2014-2017. Kemudian melanjutkan pendidikan pada Perguruan Tinggi Program Studi Sistem Informasi Fakultas Sains dan Teknologi Universitas Islam Negri Sultan Syarif Kasim Riau pada tahun 2017. Selama menjalani masa studi di UIN SUSKA Riau, penulis bergabung dengan organisasi Puzzle Research Data Teknology sebagai anggota aktif. Penulis pernah melaksanakan Kerja Praktek di Kementerian Agama Kota Pekanbaru dan Kuliah Kerja Nyata (KKN) DI Nagari Simpang Tonang, Kecamatan Dua Koto, Kabupaten Pasaman. Untuk menjalin komunikasi dengan penulis baik diluar kampus maupun didalam kampus dapat menghubungi kontak melalui handphone: 0821-7036-0014 dan e-mail : nurhalimah1499@gmail.com.

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.