

**ANALISIS SENTIMEN TERHADAP KINERJA PEMERINTAH DALAM
MENANGANI COVID-19 MENGGUNAKAN METODE NAIVE BAYES**

SKRIPSI



disusun oleh

Brayen Bakara

17.11.1539

**PROGRAM SARJANA
PROGRAM STUDI INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2021**

**ANALISIS SENTIMEN TERHADAP KINERJA PEMERINTAH DALAM
MENANGANI COVID-19 MENGGUNAKAN METODE NAIVE BAYES**

SKRIPSI

untuk memenuhi sebagian persyaratan
mencapai gelar Sarjana
pada Program Studi Informatika



disusun oleh

Brayen Bakara

17.11.1539

**PROGRAM SARJANA
PROGRAM STUDI INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2021**

PERSETUJUAN

SKRIPSI

**ANALISIS SENTIMEN TERHADAP KINERJA PEMERINTAH DALAM
MENANGANI COVID-19 MENGGUNAKAN METODE NAIVE BAYES**

yang dipersiapkan dan disusun oleh

Brayen Bakara

17.11.1539

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 11 Desember 2020

Dosen Pembimbing,

Ainul Yaqin, M. Kom.

NIK. 190302255

PENGESAHAN

SKRIPSI

**ANALISIS SENTIMEN TERHADAP KINERJA PEMERINTAH DALAM
MENANGANI COVID-19 MENGGUNAKAN METODE NAÏVE BAYES**

yang dipersiapkan dan disusun oleh

Brayen Bakara

17.11.1539

telah dipertahankan di depan Dewan Penguji
pada tanggal 20 Desember 2021

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Yoga Pristyanto, S. Kom., M. Eng
NIK. 190302412

Joko Dwi Santoso, M. Kom.
NIK. 190302181

Alnul Yaqln, M. Kom.
NIK. 190302255

Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 26 Desember 2021

DEKAN FAKULTAS ILMU KOMPUTER

Hanif Al Fatta, M. Kom.
NIK. 190302096

PERNYATAAN

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

Seperti sesuatu yang terkait dengan naskah dan karya yang telah dibuat adalah menjadi tanggungjawab saya pribadi.

Yogyakarta, 20 Desember 2021



Brayen Bakara

17.11.1539

MOTTO

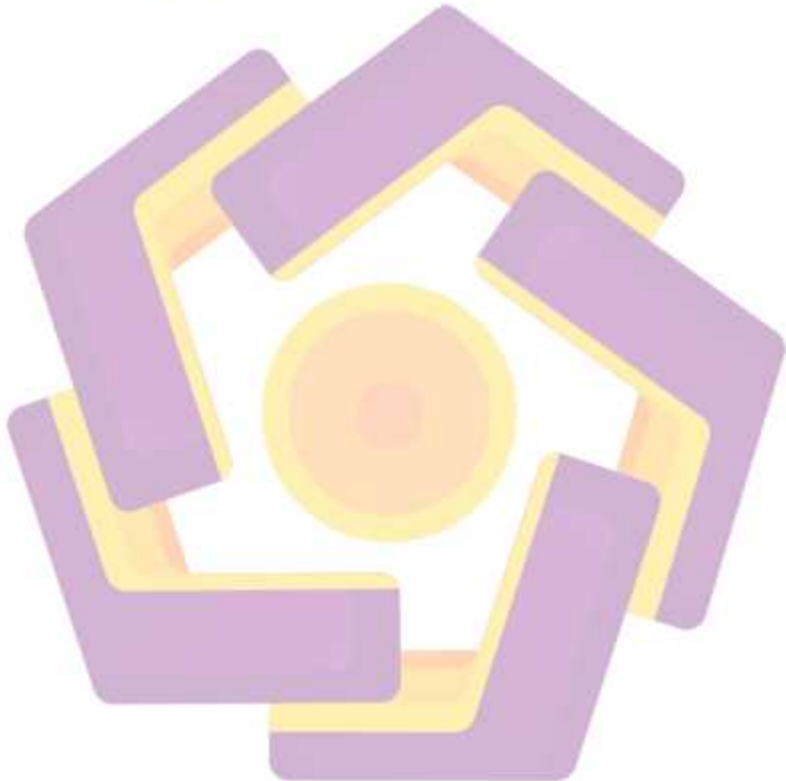
"Jangan mudah menyerah dalam segala hal. Hal memalukan bukanlah ketika kita gagal, tetapi ketika kita tidak mau mencoba lagi."

"Untuk mendapatkan kesuksesan, keberanianmu harus lebih besar dari ketakutanmu."

"Yesterday is history. Tomorrow is a mystery. Today is a gift. That's why we call it 'the present'." - Eleanor Roosevelt

PERSEMBAHAN

Skripsi ini adalah persembahan kecil saya kepada keluarga saya. Ketika dunia menutup pintunya pada saya, keluarga saya membuka lengannya untuk saya. Ketika orang-orang menutup telinga mereka untuk saya, mereka membuka hati untukku. Terima kasih karena selalu ada untukku.



KATA PENGANTAR

Puji syukur saya panjatkan kepada Tuhan Yang Maha Esa atas rahmat dan berkat-Nya sehingga penulis dapat menyelesaikan skripsi ini dengan baik dan lancar. Perlu diketahui bahwa skripsi ini merupakan salah satu syarat bagi setiap mahasiswa/i Program Studi Informatika, Fakultas Ilmu Komputer Universitas AMIKOM Yogyakarta untuk memperoleh gelar sarjana. Tidak dapat disangkal bahwa butuh usaha yang keras dalam pengerjaan skripsi ini. Namun, karya ini tidak akan selesai tanpa dukungan dari orang – orang tercinta di sekeliling saya. Terima kasih saya sampaikan kepada:

1. Bapak Ainul Yaqin, M.Kom selaku Dosen Pembimbing Skripsi saya yang dengan sabar memberi petunjuk, bimbingan dan waktunya kepada penulis dalam pembuatan skripsi ini.
2. Bapak Yoga Pristyanto, S.Kom, M.Eng dan Bapak Joko Dwi Santoso, M.Kom selaku Dosen Penguji Skripsi yang membuat pendadaran saya lebih santai dan tidak tegang.
3. Keluarga penulis yang telah memberi banyak semangat dan dukungan moral agar dapat mengerjakan skripsi ini dengan baik.

Penulis sadar bahwa masih banyak kekurangan dalam penyusunan skripsi ini oleh karena itu penulis menerima segala kritik dan saran yang bersifat membangun demi kesempurnaan skripsi ini. Akhir kata semoga skripsi ini dapat bermanfaat bagi pembaca dan dapat dijadikan referensi demi pengembangan ke arah yang lebih baik.

DAFTAR ISI

HALAMAN JUDUL.....	i
HALAMAN PERSETUJUAN.....	ii
HALAMAN PENGESAHAN.....	iii
HALAMAN PERNYATAAN.....	iv
MOTTO.....	v
PERSEMBAHAN.....	vi
KATA PENGANTAR.....	vii
DAFTAR ISI.....	viii
DAFTAR TABEL.....	xi
DAFTAR GAMBAR.....	xii
INTISARI.....	xiii
ABSTRACT.....	xiv
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang Masalah.....	1
1.2 Rumusan Masalah.....	3
1.3 Tujuan.....	3
1.4 Manfaat.....	3
1.5 Batasan Masalah.....	4
1.6 Sistematika Penulisan.....	4
BAB II LANDASAN TEORI.....	6
2.1 Tinjauan Pustaka.....	6
2.2 Natural Language Processing.....	11

2.3	Text Mining.....	11
2.4	Analisis Sentimen.....	12
2.5	Twitter.....	13
2.6	Data Preprocessing.....	13
2.7	TF-IDF.....	14
2.8	Naïve Bayes Classifier.....	15
2.9	Confusion Matrix.....	17
2.10	Flask.....	20
BAB III METODOLOGI PENELITIAN.....		21
3.1	Deskripsi Umum.....	21
3.2	Perancangan Sistem.....	22
3.2.1	Pengumpulan Data.....	22
3.2.2	Preprocessing Data.....	23
3.2.3	Pelabelan Sentimen dengan Metode Lexicon Based.....	28
3.2.4	TF-IDF.....	29
3.2.5	SMOTE.....	30
3.2.6	Implementasi Algoritma Naïve Bayes Classifier.....	31
3.2.7	Pengujian Akurasi Algoritma Naïve Bayes Classifier.....	38
3.2.8	Menyimpan dan Menerapkan Model.....	39
BAB IV HASIL DAN PEMBAHASAN.....		40
4.1	Implementasi.....	40
4.1.1	Pengumpulan Data.....	40
4.1.2	Preprocessing Data.....	41
4.1.3	Pelabelan Sentimen dengan Textblob.....	46
4.1.4	TF-IDF.....	48

4.1.5	Resampling Menggunakan SMOTE	50
4.1.6	Implementasi Algoritma Naïve Bayes Classifier	50
4.1.7	Hasil Klasifikasi Sentimen	51
4.1.8	Menyimpan dan Menerapkan Model	52
4.2	Hasil dan Pembahasan	53
4.2.1	Hasil Klasifikasi Sentimen	53
4.2.2	Hasil Akurasi, Presisi dan Recall	54
4.3	User Interface	58
BAB V PENUTUP		61
5.1	Kesimpulan	61
5.2	Saran	62
DAFTAR PUSTAKA		63

DAFTAR TABEL

Tabel 2.1 Perbandingan Penelitian Penulis dan Penelitian Terdahulu.....	7
Tabel 2.2 <i>Confusion Matrix</i> dengan Kelas 3x3.....	18
Tabel 3.1 Sampel Data Tweet.....	22
Tabel 3.2 Contoh Proses Case Folding.....	24
Tabel 3.3 Contoh Proses Cleansing.....	24
Tabel 3.4 Contoh Proses Tokenizing.....	25
Tabel 3.5 Contoh Proses Normalization.....	25
Tabel 3.6 Contoh Proses Remove Stop Word.....	26
Tabel 3.7 Contoh Proses Stemming.....	28
Tabel 3.8 Contoh Data Tweets yang Sudah Dilabeli.....	28
Tabel 3.9 Contoh Data Tweets.....	29
Tabel 3.10 Contoh Perhitungan TF-IDF.....	30
Tabel 3.11 Contoh Kasus Data Training.....	32
Tabel 3.12 Contoh Perhitungan Prior Probabilitas untuk Tiap Kelas.....	33
Tabel 3.13 Contoh Perhitungan Conditional Probability untuk Tiap Kelas.....	33
Tabel 3.14 Contoh Data Testing.....	35
Tabel 3.15 Hasil Proses Matching Data Testing dan Data Training.....	36
Tabel 3.16 Hasil Perhitungan Posterior Probability untuk Tiap Kelas.....	37
Tabel 4.1 Data untuk Sentimen Analisis.....	53
Tabel 4.2 Hasil Confusion Matrix NBC tanpa SMOTE.....	55
Tabel 4.3 Hasil Confusion Matrix NBC dengan SMOTE.....	55

DAFTAR GAMBAR

Gambar 3.1	Gambaran Umum Sistem	21
Gambar 4.1	Kumpulan Data Tweet	41
Gambar 4.2	Fungsi Proses Case Folding	42
Gambar 4.3	Fungsi Proses Cleansing.....	42
Gambar 4.4	Fungsi Proses Tokenizing	43
Gambar 4.5	Fungsi Proses Normalization	43
Gambar 4.6	Fungsi Proses Remove Stop Word.....	44
Gambar 4.7	Fungsi Proses Stemming	45
Gambar 4.8	Fungsi Proses Penerjemahan.....	46
Gambar 4.9	Fungsi Proses Pelabelan.....	47
Gambar 4.10	Hasil Pelabelan dengan Textblob.....	48
Gambar 4.11	Proses TF-IDF.....	49
Gambar 4.12	Proses Resampling Menggunakan SMOTE.....	50
Gambar 4.13	Proses Implementasi Algoritma Naïve Bayes Classifier.....	50
Gambar 4.14	Proses Hasil Klasifikasi Sentimen.....	51
Gambar 4.15	Proses Menyimpan dan Menerapkan Model.....	52
Gambar 4.16	Hasil Klasifikasi Sentimen	54
Gambar 4.17	Hasil Akurasi Klasifikasi Sentimen	56
Gambar 4.18	Hasil Presisi, Recall, dan F1-Score Klasifikasi tanpa SMOTE.....	56
Gambar 4.19	Hasil Presisi, Recall, dan F1-Score Klasifikasi dengan SMOTE....	57
Gambar 4.20	Tampilan Bagian Preprocessing Data	58
Gambar 4.21	Tampilan Bagian Klasifikasi Lexicon.....	58
Gambar 4.22	Tampilan Hasil Klasifikasi NBC tanpa SMOTE	59
Gambar 4.23	Tampilan Hasil Klasifikasi NBC tanpa SMOTE	59
Gambar 4.24	Tampilan Halaman Klasifikasi Teks	60
Gambar 4.25	Tampilan Halaman Hasil Klasifikasi Teks.....	60

INTISARI

Dewasa ini, penggunaan internet sudah menjadi bagian dari kehidupan masyarakat terutama penggunaan media social seperti *twitter*. Sumber data yang jumlahnya banyak ini dapat digunakan untuk berbagai tujuan. Contohnya membantu pemerintah dalam membuat kebijakan dalam penanganan pandemic. Analisis sentimen menggunakan algoritma Naïve Bayes dapat dipilih untuk mengelola sumber data tersebut. Dalam penelitian ini, peneliti mencoba untuk membuat analisis sentimen untuk mengelompokkan tanggapan masyarakat terhadap penanganan COVID-19 oleh pemerintah dan mengetahui tingkat akurasi pengelompokkannya.

Tahapan proses dimulai dari pengumpulan data dan dilanjutkan dengan proses *preprocessing data*. Proses ini terdiri dari *case folding*, *cleansing*, *tokenizing*, *normalization*, *remove stop word*, dan *stemming*. Data yang sudah diproses kemudian akan dilabeli dengan metode *lexicon based* lalu dibagi menjadi *data training* dan *data testing*. Setelah itu barulah data dilakukan proses pembobotan menggunakan TF-IDF. Di sini *data training* akan diproses menggunakan *Synthetic Minority Over Sampling Technique* akibat data yang tidak seimbang. Selanjutnya *data training* akan digunakan sebagai bahan latih algoritma Naïve Bayes Classifier dan diuji menggunakan *data testing*. Untuk melihat kemampuan system digunakan *confusion matrix*.

Dari penelitian yang telah dilakukan, diperoleh akurasi pemodelan Naïve Bayes Classifier tanpa *Synthetic Minority Over Sampling Technique* sebesar 68,88% sedangkan pada pemodelan dengan *Synthetic Minority Over Sampling Technique* diperoleh akurasi sebesar 66,71%. Meskipun begitu, terjadi peningkatan nilai *precision*, *recall*, dan pada kelas minoritas. Pada kelas netral, nilai *precision* tanpa *Synthetic Minority Over Sampling Technique* bernilai 39%, sedangkan dengan *Synthetic Minority Over Sampling Technique* bernilai 70%. Untuk *recall*, nilainya sebelum *Synthetic Minority Over Sampling Technique* sebesar 49% sedangkan dengan *Synthetic Minority Over Sampling Technique* menjadi 64%.

Kata kunci: *Tweet, TF-IDF, Synthetic Minority Over Sampling Technique, Naïve Bayes Classifier, Lexicon Based, Confusion Matrix*

ABSTRACT

Today, the use of the internet has become a part of people's lives, especially the use of social media such as Twitter. This large number of data sources can be used for a variety of purposes. For example, helping the government in making policies in handling pandemics. Sentiment analysis using the Naïve Bayes algorithm can be selected to manage the data source. In this study, researcher tried to create a sentiment analysis to group the response to the government's handling of COVID-19 and find out the accuracy of the grouping.

The process stage starts from data collection and continues with the process of preprocessing data. This process consists of case folding, cleansing, tokenizing, normalization, remove stop word, and stemming. The data that has been processed will then be labeled with a lexicon based method and then divided into training data and testing data. After that, the data was carried out the weighting process using TF-IDF. Here the training data will be processed using synthetic minority over sampling technique due to unbalanced data. Furthermore, the training data will be used as a training material for the Naïve Bayes Classifier algorithm and tested using testing data. To see the system capabilities confusion matrix will be used.

From the research that has been done, obtained the accuracy of naïve bayes classifier modeling without Synthetic Minority Over Sampling Technique by 68.88% while in modeling with Synthetic Minority Over Sampling Technique obtained accuracy of 66.71%. Even so, there was an increase in precision value, recall, and in the minority class. In neutral classes, precision value without Synthetic Minority Over Sampling Technique is 39%, while synthetic minority over sampling technique is worth 70%. For recall, the value before Synthetic Minority Over Sampling Technique is 49% while with Synthetic Minority Over Sampling Technique is up to 64%.

Keyword: Tweet, TF-IDF, Synthetic Minority Over Sampling Technique, Naïve Bayes Classifier, Lexicon Based, Confusion Matrix