

**ANALISIS SENTIMENT PENGGUNA TWITTER TERHADAP
PEMBELAJARAN JARAK JAUH PADA MASA
COVID19 MENGGUNAKAN ALGORITMA
NAIVE BAYES CLASSIFIER**

SKRIPSI



disusun oleh

Adi Kresna Kencana

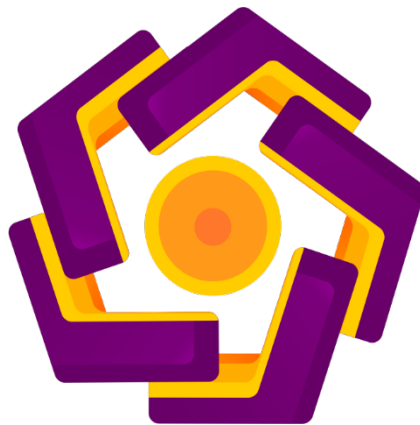
17.11.1515

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

**ANALISIS SENTIMENT PENGGUNA TWITTER TERHADAP
PEMBELAJARAN JARAK JAUH PADA MASA
COVID19 MENGGUNAKAN ALGORITMA
NAIVE BAYES CLASSIFIER**

SKRIPSI

untuk memenuhi sebagian persyaratan
mencapai gelar Sarjana
pada Program Studi Informatika



disusun oleh

Adi Kresna Kencana

17.11.1515

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

PERSETUJUAN

SKRIPSI

**ANALISIS SENTIMENT PENGGUNA TWITTER TERHADAP
PEMBELAJARAN JARAK JAUH PADA MASA
COVID19 MENGGUNAKAN METODE
NAIVE BAYES CLASSIFIER**

yang dipersiapkan dan disusun oleh

Adi Kresna Kencana

17.11.1515

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 08 Februari 2021

Dosen Pembimbing,

Arif Akbarul Huda, S.Si, M.Eng.
NIK. 190302287

PENGESAHAN

SKRIPSI

**ANALISIS SENTIMENT PENGGUNA TWITTER TERHADAP
PEMBELAJARAN JARAK JAUH PADA MASA
COVID19 MENGGUNAKAN METODE
NAIVE BAYES CLASSIFIER**

yang dipersiapkan dan disusun oleh

Adi Kresna Kencana

17.11.1515

telah dipertahankan di depan Dewan Penguji
pada tanggal 18 Februari 2021

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Arif Akbarul Huda, S.Si, M.Eng
NIK. 190302287

Bayu Setiaji, M.Kom
NIK. 190302216

Yuli Astuti, M.Kom
NIK. 190302146

Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 11 Maret 2021

**DEKAN FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA**

Krisnawati, S.Si M.M.
NIK. 190302098

PERNYATAAN

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

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

Yogyakarta, ... MARET 2021

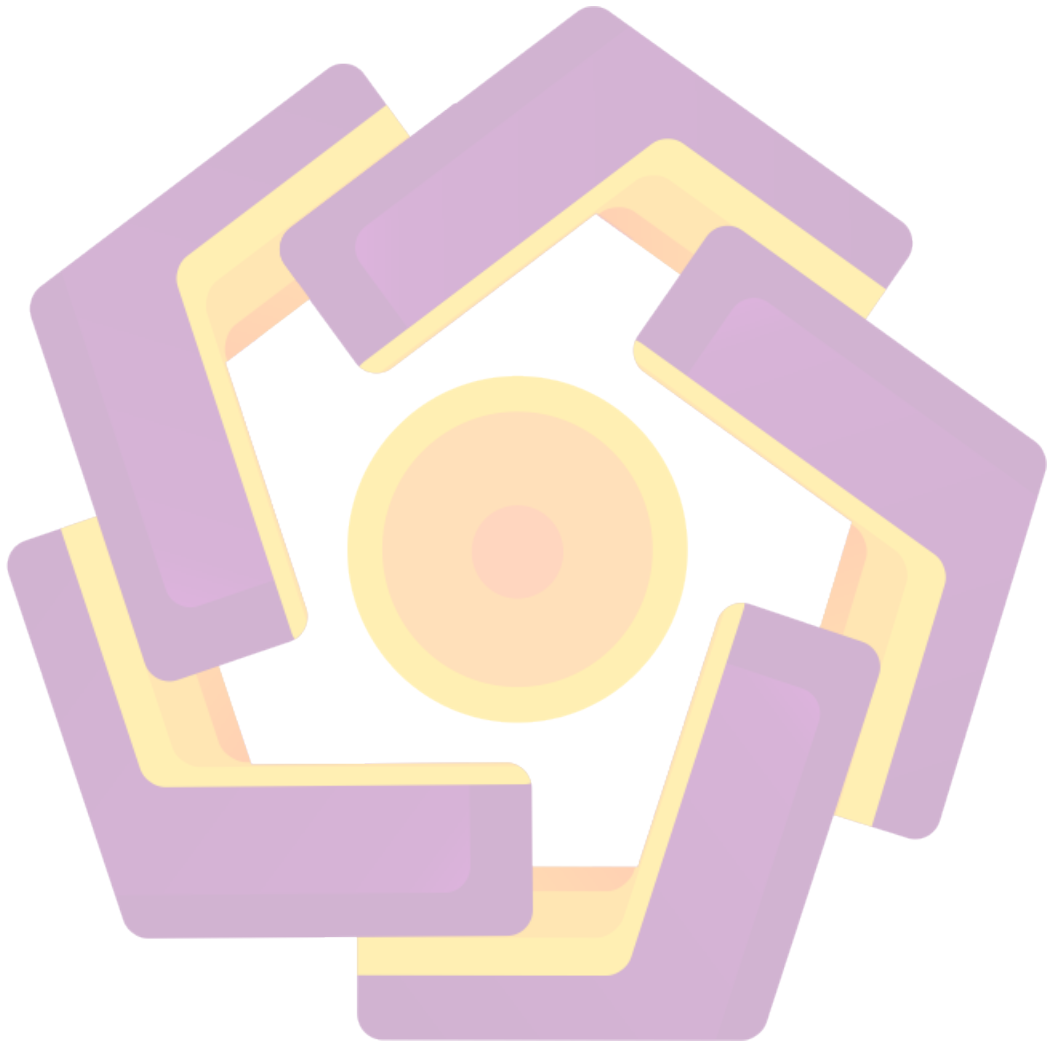


Adi Kresna Kencana

NIM. 17.11.1515

MOTTO

"Saya harus selalu ingat bahwa entropi tidak ada dipihakmu"



PERSEMBAHAN

Skripsi ini di persembahkan kepada:

1. Orang tua dan adik-adik saya, Ibu Suryani dan Bapak Mujiyono yang selalu mendoakan setiap saat dan memberikan semangat tanpa henti serta adik saya Fathan dan Iza yang selalu mendoakan kakaknya.
2. Kepada Bapak Arif Akbarul Huda, S.Si, M.Eng yang telah membimbing saya dari kesusahan dalam mengerjakan skripsi dan memberikan masukan-masukan untuk membantu saya dalam mempercepat menyelesaikan skripsi ini.
3. Semua Dosen S1 Informatika atas ilmu yang diberikan.
4. Teman satu kosan dan seperjuangan IF-09 semoga tetap kompak dan semangat.

KATA PENGANTAR

Puji syukur kehadiran Allah SWT atas berkat dan rahmatnya akhirnya dapat menyelesaikan skripsi ini. Skripsi ini dibuat untuk memenuhi syarat menyelesaikan studi Informatika jenjang Strata 1 Universitas Amikom Yogyakarta.

Penulis sadar bahwa dalam menyelesaikan skripsi ini tidak lepas dari bantuan berbagai pihak untuk memberikan bantuan, bimbingan, bantuan dalam membuat teks secara baik. Dengan kesempatan ini penulis mengucapkan terima kasih sebesar-besarnya kepada:

1. Bapak Arif Akbarul Huda, S.Si, M.Eng. selaku Dosen yang membimbing saya dari kesusahan dalam mengerjakan judul pertama akhirnya di bantu dan merubah judul, akhirnya selesai juga terima kasih Pak Arif.
2. Ibu, Bapak, dan adik saya yang telah memberikan banyak doa dan memotivasi agar cepat menyelesaikan skripsi ini.
3. Terima kasih untuk seluruh Dosen Program Studi Informatika Amikom Yogyakarta yang telah memberikan ilmunya.
4. Terima kasih untuk Teman satu kosan yang sudah sangat membantu dalam menyelesaikan skripsi ini serta teman-teman IF-09.

Terima kasih semua yang sudah membantu semoga segala bantuan yang diberikan sehingga dapat menyelesaikan skripsi ini dicatat oleh Allah SWT.

Yogyakarta, 11 Maret 2021

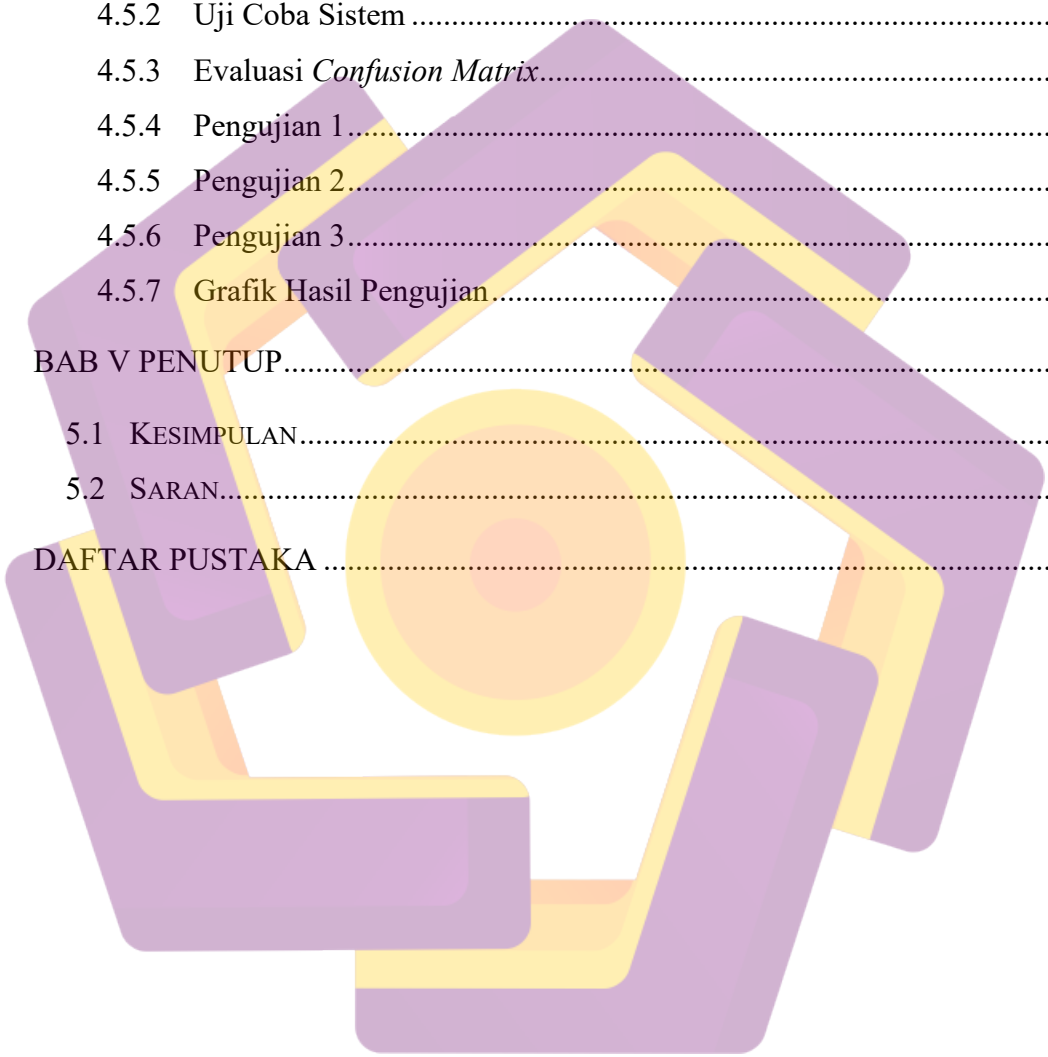


Adi Kresna Kencana

DAFTAR ISI

JUDUL.....	I
PERSETUJUAN	II
PENGESAHAN	III
PERNYATAAN.....	V
MOTTO	VI
PERSEMBAHAN.....	VII
KATA PENGANTAR	VIII
DAFTAR ISI.....	IX
DAFTAR TABEL.....	XII
DAFTAR GAMBAR	XIII
DAFTAR SOURCE CODE.....	XIV
INTISARI.....	XV
<i>ABSTRACT</i>	XVI
BAB I PENDAHULUAN.....	1
1.1 LATAR BELAKANG.....	1
1.2 RUMUSAN MASALAH	2
1.3 BATASAN MASALAH.....	3
1.4 MAKSUD DAN TUJUAN PENELITIAN	3
1.5 MANFAAT PENELITIAN.....	4
1.6 METODE PENELITIAN	4
1.6.1 Metode Pengumpulan Data.....	4
1.6.2 Analisis Permasalahan	4
1.6.3 Metode Perancangan	4
1.7 SISTEMATIKA PENULISAN	5
BAB II LANDASAN TEORI.....	6

2.1	TINJAUAN PUSTAKA.....	6
2.2	SENTIMEN ANALISIS	6
2.3	<i>TEXT MINING</i>	7
2.4	<i>TEXT PREPROCESSING</i>	7
2.5	PEMBOBOTAN TF-IDF	8
2.6	<i>NAIVE BAYES CLASSIFIER</i>	9
2.6.1	<i>Multinomial Naive Bayes Classifier</i>	10
2.7	EVALUASI	10
BAB III METODE PENELITIAN.....		12
3.1	ALUR PENELITIAN.....	12
3.2	PERANCANGAN DESAIN SISTEM	13
3.3	ALAT DAN BAHAN PENELITIAN.....	13
3.3.1	Alat Penelitian.....	13
3.3.2	Bahan Penelitian.....	14
3.4	<i>PREPROCESSING DATA</i>	16
3.4.1	<i>Case Folding</i>	16
3.4.2	<i>Tokenizing</i>	17
3.4.3	<i>Stopword Removal</i>	18
3.4.5	<i>Normalisasi</i>	18
3.4.6	<i>Stemming</i>	19
3.5	PEMBOBOTAN TF-IDF	19
3.6	PROSES KLASIFIKASI.....	20
3.7	EVALUASI	21
BAB IV HASIL DAN PEMBAHASAN		23
4.1	PENGUMPULAN DATA	23
4.2	<i>PREPROCESSING DATA</i>	24
4.2.1	<i>Case Folding</i>	25
4.2.2	<i>Tokenizing</i>	25
4.2.3	<i>Stopword Removal</i>	27
4.2.4	<i>Normalisasi</i>	28



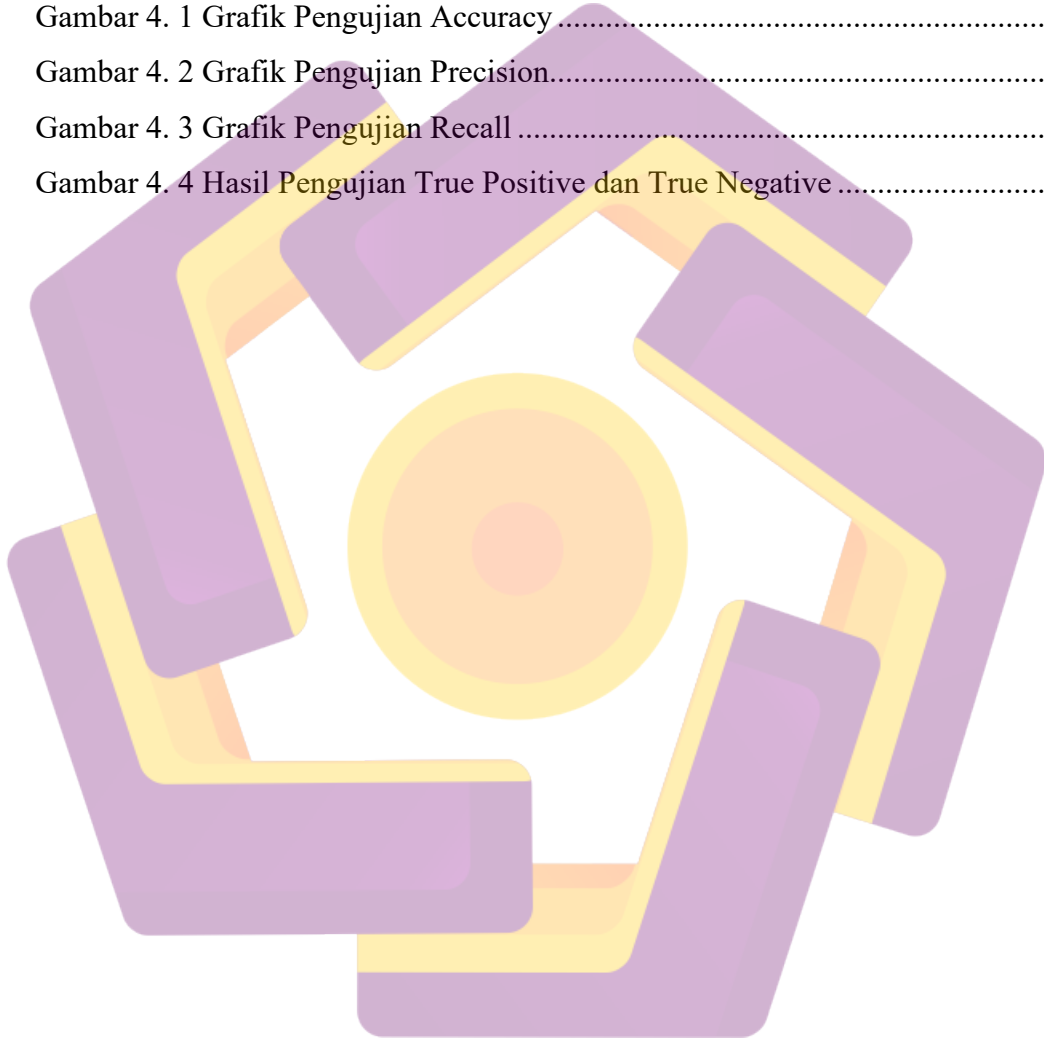
4.2.5	<i>Stemming</i>	29
4.3	PEMBAGIAN DATA	30
4.4	PEMBOBOTAN TF-IDF	31
4.5	IMPLEMENTASI KLASIFIKASI	31
4.5.1	Klasifikasi <i>Naïve Bayes Classifier</i>	31
4.5.2	Uji Coba Sistem	32
4.5.3	Evaluasi <i>Confusion Matrix</i>	32
4.5.4	Pengujian 1	33
4.5.5	Pengujian 2	34
4.5.6	Pengujian 3	35
4.5.7	Grafik Hasil Pengujian	36
BAB V PENUTUP		39
5.1	KESIMPULAN	39
5.2	SARAN	39
DAFTAR PUSTAKA		40

DAFTAR TABEL

Tabel 2. 1 Confusion Matrix	11
Tabel 3. 1 Dokumen Tweet Teks	15
Tabel 3. 2 Proses Case Folding.....	17
Tabel 3. 3 Proses Tokenizing.....	17
Tabel 3. 4 Proses Stopword Removal	18
Tabel 3. 5 Proses Normalisasi.....	18
Tabel 3. 6 Proses Stemming.....	19
Tabel 3. 7 Proses TF-IDF	20
Tabel 3. 8 Confusion Matrix	21
Tabel 4. 1 Data Label.....	24
Tabel 4. 2 Case Folding	25
Tabel 4. 3 Tokenizing	27
Tabel 4. 4 Stopword Removal.....	28
Tabel 4. 5 Normalisasi	29
Tabel 4. 6 Stemming	30
Tabel 4. 7 Data Training dan Testing.....	32
Tabel 4. 8 Hasil Confusion Matrix Pengujian 1.....	33
Tabel 4. 9 Hasil Confusion Matrix Pengujian 2.....	34
Tabel 4. 10 Hasil Confusion Matrix Pengujian 3.....	35

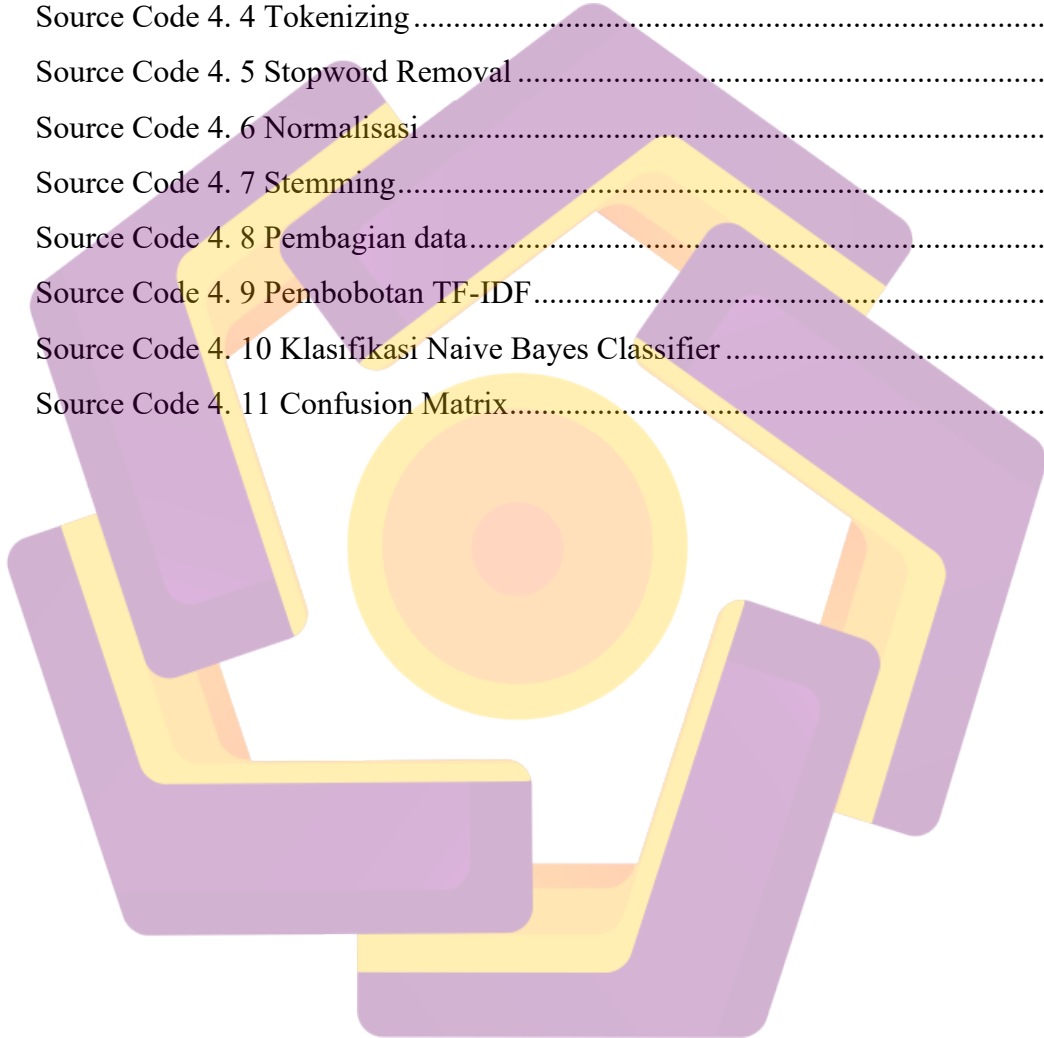
DAFTAR GAMBAR

Gambar 3. 1 Alur Penelitian.....	12
Gambar 3. 2 Crawling Data Twitter.....	15
Gambar 3. 3 Proses Preprocessing	16
Gambar 4. 1 Grafik Pengujian Accuracy.....	37
Gambar 4. 2 Grafik Pengujian Precision.....	37
Gambar 4. 3 Grafik Pengujian Recall	38
Gambar 4. 4 Hasil Pengujian True Positive dan True Negative	38



DAFTAR SOURCE CODE

Source Code 4. 1 Api Token.....	23
Source Code 4. 2 Pengambilan Dataset	24
Source Code 4. 3 Case Folding.....	25
Source Code 4. 4 Tokenizing.....	26
Source Code 4. 5 Stopword Removal	28
Source Code 4. 6 Normalisasi.....	29
Source Code 4. 7 Stemming.....	30
Source Code 4. 8 Pembagian data.....	30
Source Code 4. 9 Pembobotan TF-IDF.....	31
Source Code 4. 10 Klasifikasi Naive Bayes Classifier	32
Source Code 4. 11 Confusion Matrix.....	33



INTISARI

Pada masa *pandemic* Covid-19 pemerintah Indonesia melalui Menteri Pendidikan dan Kebudayaan Republik Indonesia memberikan kebijakan baru dalam Pendidikan yaitu pembelajaran jarak jauh. Setelah kebijakan baru tersebut dilaksanakan banyak memberikan pengaruh pada masyarakat. Banyak dari masyarakat memberikan opini mereka di sosial media terutama twitter, dari opini-opini masyarakat terdapat *tweet* pada twitter mengandung kata positif maupun negatif terhadap pembelajaran jarak jauh. Dari permasalahan ini dilakukan penelitian menggunakan analisis sentiment.

Untuk mendapatkan evaluasi hasil yang terbaik dilakukan perbandingan data latih dan data uji. Tahapan penelitian yang dilakukan terdiri dari pengambilan data dari media sosial twitter, *preprocessing data*, pembobotan data TF-IDF dan klasifikasi menggunakan algoritma *Naïve Bayes Classifier*.

Dari data yang sudah dibersihkan diberikan label dengan klasifikasi secara manual, terdapat dua label yaitu positif dan negatif. Setelah dataset diberikan label secara manual selanjutnya pembagian data. Pada pembagian data data terpisah menjadi dua kategori yaitu kategori *training* dan *testing*, pada dataset jumlah dibagi menjadi 3 dataset antara lain 600 dataset, 500 dataset dan 400 dataset. Setelah dilakukan pembagian data dilakukan pengujian dengan *confusion matrix*, dilakukan pengujian sebanyak 3 percobaan pengujian dengan hasil akurasi tertinggi yang didapatkan *naïve bayes classifier* dengan model *multinomial* sebesar 83%.

Kata Kunci: *Naïve Bayes Classifier*, Sentimen Analisis, Klasifikasi, *Text Mining*

ABSTRACT

During the Covid-19 pandemic, the Indonesian government through the Minister of Education and Culture of the Republic of Indonesia provided a new policy in education, namely distance learning. After the new policy was implemented, it had a lot of influence on society. Many people give their opinion on social media, especially twitter, from the public's opinions there are tweets on Twitter that contain both positive and negative words about distance learning. From this problem, a study was conducted using sentiment analysis.

To get the best evaluation of results, the comparison of training data and test data is carried out. The research stages consisted of data collection from social media twitter, data preprocessing, weighting of TF-IDF data and classification using the Naïve Bayes Classifier algorithm.

From the data that has been cleaned, it is labeled with manual classification, there are two labels, positive and negative. After the dataset is labeled manually, then the data is shared. In the division of data, the data is separated into two categories, namely the training and testing category, in the dataset the number is divided into 3 datasets, including 600 datasets, 500 datasets and 400 datasets. After sharing the data, the test was carried out with confusion matrix. Three testing experiments were carried out with the highest accuracy results obtained by the naïve Bayes classifier with a multinomial model of 83%.

Keyword: *Naïve Bayes Classifier, Sentiment Analysis, Classification, Text Mining*