

**ANALISIS SENTIMENT PENGGUNA TWITTER TERHADAP
PROVIDER BY.U MENGGUNAKAN ALGORITMA
NAIVE BAYES CLASSIFIER**

SKRIPSI



disusun oleh

Bagas Sonas Audit

17.11.1490

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

**ANALISIS SENTIMENT PENGGUNA TWITTER TERHADAP
PROVIDER BY.U MENGGUNAKAN ALGORITMA
NAIVE BAYES CLASSIFIER**

SKRIPSI

untuk memenuhi sebagian persyaratan
mencapai gelar Sarjana
pada Program Studi Informatika



disusun oleh
Bagas Sonas Audit
17.11.1490

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

PERSETUJUAN

SKRIPSI

**ANALISIS SENTIMENT PENGGUNA TWITTER TERHADAP
PROVIDER BY.U MENGGUNAKAN ALGORITMA
NAIVE BAYES CLASSIFIER**

yang dipersiapkan dan disusun oleh

Bagas Sonas Audit

17.11.1490

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 03 Juni 2021

Dosen Pembimbing,

Ike Verawati, M.Kom

NIK. 190302237

PENGESAHAN

SKRIPSI

ANALISIS SENTIMENT PENGGUNA TWITTER TERHADAP PROVIDER BY.U MENGGUNAKAN ALGORITMA NAIVE BAYES CLASSIFIER

yang dipersiapkan dan disusun oleh

Bagas Sonas Audit

17.11.1490

telah dipertahankan di depan Dewan Penguji
pada tanggal 26 Juli 2021

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Achmah Sidauruk, M.Kom
NIK. 190302238

Sri Ngudi Wahyuni, S.T., M.Kom.
NIK. 190302060

Ike Verawati, M. Kom
NIK. 190302237

Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 27 Juli 2021

DEKAN FAKULTAS ILMU KOMPUTER

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

PERNYATAAN

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

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

Yogyakarta, 20 Juli 2021

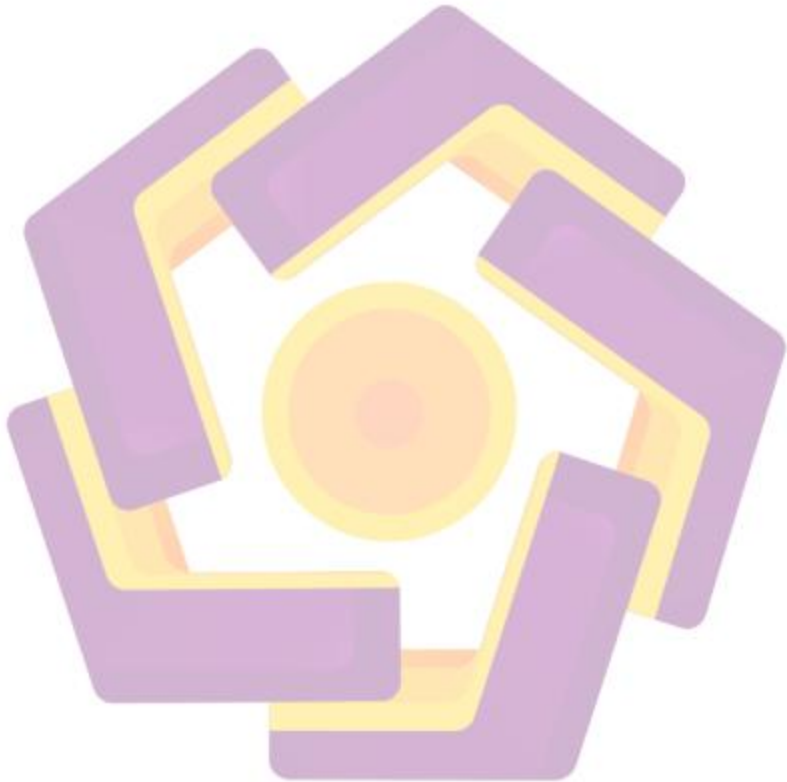


Bagas Sonas Audit

NIM. 17.11.1490

MOTTO

"Ambilah kebaikan dari apa yang dikatakan, jangan melihat siapa yang mengatakannya"



PERSEMBAHAN

Skripsi ini di persembahkan kepada:

1. Kedua orang tua, Bapak Wahyu Indarto dan Ibu Ratnawati yang tidak lelah berdoa untuk memberikan semangat, serta adik-adik saya Nanang dan Fajar.
2. Kepada Ibu Ike Verawati, M. Kom yang telah membimbing saya dari kesusahan dalam mengerjakan skripsi dan memberikan saran untuk mempercepat saya menyelesaikan skripsi.
3. Semua Dosen SI Informatika terimakasih untuk ilmu yang sudah diberikan.
4. Teman berjuang dari semester awal Bima, Restu, Dandhi, dan Sungsang. Teman kosan Adi, Andhi, Bagas, Farhan, dan Fadhil. Serta teman-teman satu kelas perjuangan IF-09 semoga tetap semangat.

KATA PENGANTAR

Puji syukur kehadiran Tuhan Yang Maha Esa atas berkat, rahmat dan petunjuknya sehingga saya dapat menyelesaikan skripsi ini dengan baik. Skripsi ini telah dibuat untuk memenuhi syarat menyelesaikan studi Informatika jenjang Strata I di Universitas Amikom Yogyakarta.

Dalam menyelesaikan penyusunan skripsi ini tidak lepas dari bantuan, arahan, serta dukungan dari berbagai pihak. Dengan kesempatan ini penulis mengucapkan terima kasih kepada:

1. Kedua orang tua, Bapak Wahyu Indarto dan Ibu Ratnawati, kakung Toha Dasiran, serta kedua adik saya Nanang Kristyanto dan Fajar Satrio Pamungkas.
2. Ibu Ike Verawati, M. Kom., selaku Dosen Pembimbing yang telah meluangkan waktu dan pikiran untuk membimbing penulis dalam menyelesaikan skripsi ini.
3. Bapak dan Ibu dosen S1 Informatika yang telah memberikan ilmu kepada penulis.
4. Teman – teman dekat serta teman kelas IF-09 yang sudah membantu dalam menyelesaikan skripsi ini.

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

Yogyakarta, 20 Juli 2021

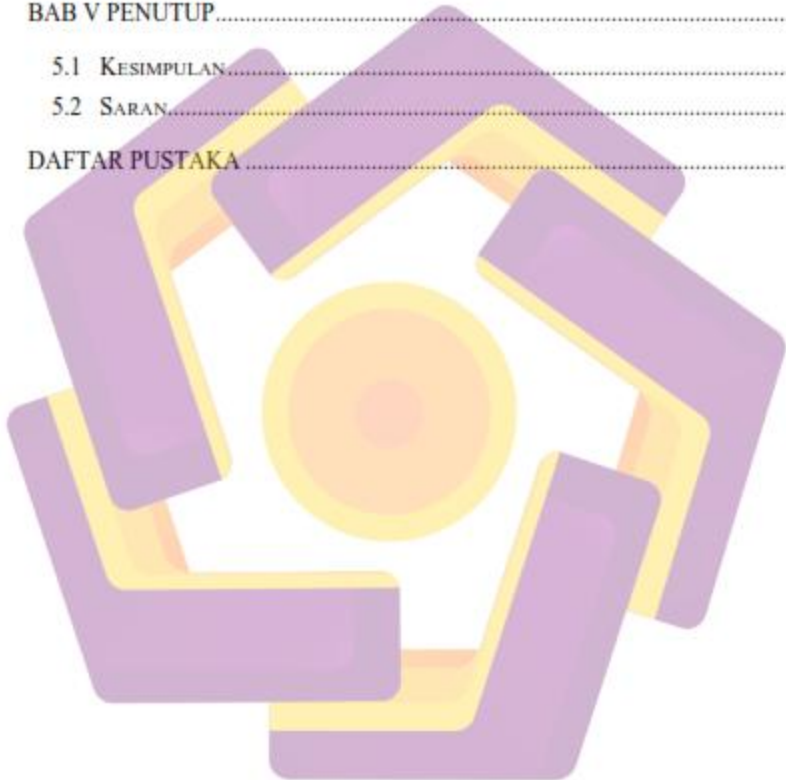
Bagas Sonas Audit

DAFTAR ISI

JUDUL.....	I
PERSETUJUAN	III
PENGESAHAN	IV
PERNYATAAN.....	V
MOTTO	VI
PERSEMBAHAN.....	VII
KATA PENGANTAR.....	VIII
DAFTAR ISI.....	IX
DAFTAR TABEL.....	XII
DAFTAR GAMBAR.....	XIII
DAFTAR SOURCE CODE.....	XIV
INTISARI.....	XV
ABSTRACT.....	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 METODOLOGI PENELITIAN	4
1.6.1 Metode Pengumpulan Data.....	4
1.6.2 Analisis Permasalahan	4
1.6.3 Metode Testing.....	5
1.7 SISTEMATIKA PENULISAN	5
BAB II LANDASAN TEORI.....	7

2.1	TINJAUAN PUSTAKA.....	7
2.2	ANALISIS SENTIMEN	9
2.3	<i>TEXT MINING</i>	10
2.4	<i>TEXT PREPROCESSING</i>	10
2.5	PEMBOBOTAN TF-IDF	11
2.6	<i>NAIVE BAYES CLASSIFIER</i>	12
2.6.1	<i>Multinomial Naïve Bayes Classifier</i>	13
2.7	EVALUASI	13
BAB III METODE PENELITIAN.....		15
3.1	ALUR PENELITIAN.....	15
3.2	ALAT DAN BAHAN PENELITIAN.....	16
3.2.1	Alat Penelitian.....	16
3.2.2	Bahan Penelitian.....	17
3.3	MENERJEMAHKAN EMOTICON	18
3.4	PENGLASIFIKASIAN NAIVE BAYES CLASSIFIER.....	18
BAB IV HASIL DAN PEMBAHASAN		21
4.1	PENGAMBILAN DATA TWITTER.....	21
4.2	<i>PREPROCESSING DATA</i>	23
4.2.1	<i>Menerjemahkan Emoticon</i>	23
4.2.2	<i>Case Folding</i>	24
4.2.3	<i>Tokenizing</i>	25
4.2.4	<i>Stopword Removal</i>	27
4.2.5	<i>Normalisasi</i>	28
4.2.6	<i>Stemming</i>	30
4.3	PROSES KLASIFIKASI/ PELABELAN.....	31
4.3.1	<i>TF-IDF</i>	32
4.3.2	<i>Klasifikasi Naïve Bayes Classifier</i>	33
4.3.3	<i>Proses Klasifikasi Otomatis</i>	34
4.4	UJI COBA SISTEM.....	35
4.5	EVALUASI	35

4.5.1	Confusion Matrix	35
4.5.2	Pengujian 1	37
4.5.3	Pengujian 2.....	38
4.5.4	Pengujian 3.....	39
4.5.5	Tabel dan Grafik Hasil Pengujian.....	40
BAB V PENUTUP.....		43
5.1	KESIMPULAN.....	43
5.2	SARAN.....	44
DAFTAR PUSTAKA.....		45

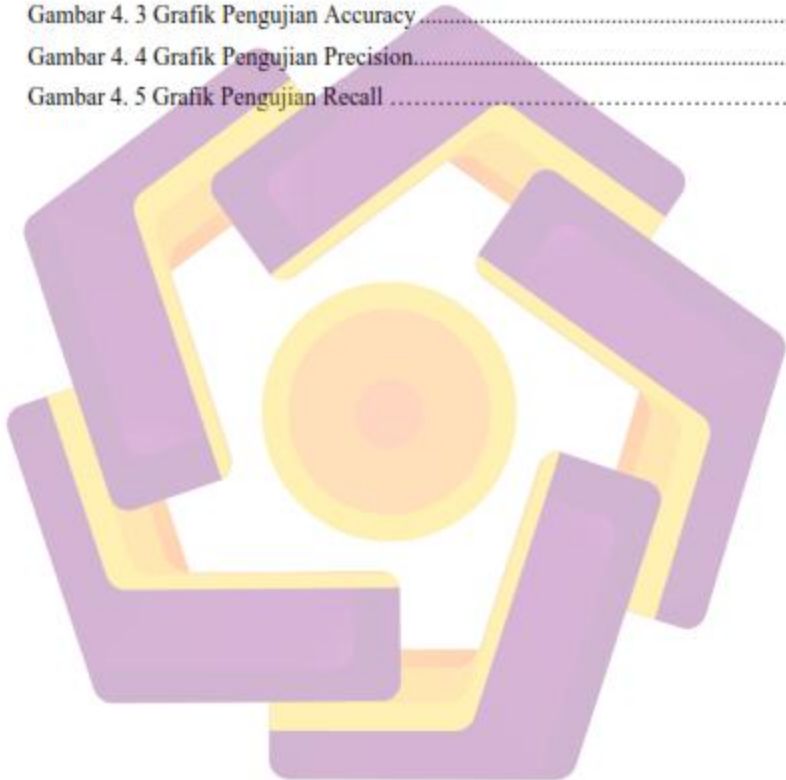


DAFTAR TABEL

Tabel 2. 1 Perbandingan Penelitian.....	7
Tabel 2. 2 Confusion Matrix	14
Tabel 3. 1 Dokumen Teks Tweet.....	17
Tabel 3. 2 Hasil Terjemahan Emoticon.....	18
Tabel 3. 3 Data Klasifikasi Otomatis.....	20
Tabel 4. 1 Data Hasil Crawling.....	22
Tabel 4. 2 Translate Emoticon	24
Tabel 4. 3 Case Folding	25
Tabel 4. 4 Tokenizing	26
Tabel 4. 5 Stopword Removal.....	28
Tabel 4. 6 Normalisasi	29
Tabel 4. 7 Stemming	30
Tabel 4. 8 Dataset Model	31
Tabel 4. 9 Proses TF-IDF.....	32
Tabel 4. 10 Hasil Klasifikasi Otomatis	34
Tabel 4. 11 Data Uji.....	35
Tabel 4. 12 Confusion Matrix	35
Tabel 4. 13 Hasil Confusion Matrix pengujian 1.....	37
Tabel 4. 14 Hasil Confusion Matrix pengujian 2.....	38
Tabel 4. 15 Hasil Confusion Matrix pengujian 3.....	39
Tabel 4. 16 Hasil Akurasi	40
Tabel 4. 17 Hasil Precision	41
Tabel 4. 18 Hasil Recall.....	41

DAFTAR GAMBAR

Gambar 3. 1 Alur Penelitian.....	15
Gambar 4. 1 Crawling Data Dari Twitter	21
Gambar 4. 2 Proses Preprocessing.....	23
Gambar 4. 3 Grafik Pengujian Accuracy.....	40
Gambar 4. 4 Grafik Pengujian Precision.....	41
Gambar 4. 5 Grafik Pengujian Recall	42



DAFTAR SOURCE CODE

Source Code 4. 1 Token API.....	21
Source Code 4. 2 Pengambilan Data.....	22
Source Code 4. 3 Translate Emoticon.....	23
Source Code 4. 4 Case Folding.....	24
Source Code 4. 5 Tokenizing.....	25
Source Code 4. 6 Stopword Removal.....	27
Source Code 4. 7 Normalisasi.....	29
Source Code 4. 8 Stemming.....	30
Source Code 4. 9 Modeling.....	31
Source Code 4. 10 TF-IDF.....	33
Source Code 4. 11 Klasifikasi.....	33
Source Code 4. 12 Klasifikas Otomatisi.....	34
Source Code 4. 13 Confusion Matrix.....	36

INTISARI

Pada beberapa tahun belakangan ini banyak bermunculan provider-provider internet baru yang memikat banyak pengguna internet di Indonesia, salah satunya yaitu by.u. Banyak dari pengguna memberikan opini mereka melalui sosial media terutama twitter, dari beragam opini pengguna terdapat *tweet* pada twitter mengandung kata-kata positif dan negatif terhadap provider by.u. Dari permasalahan ini dilakukan penelitian menggunakan analisis sentimen. Untuk mendapatkan evaluasi hasil yang terbaik dilakukan perbandingan data latih dan data uji. Tahapan penelitian yang dilakukan terdiri dari pengambilan data dari sosial media twitter, preprocessing data, pembobotan data TF-IDF dan klasifikasi menggunakan algoritma *Naïve Bayes Classifier*.

Pada penelitian ini dilakukan analisis sentimen terhadap provider by.u dengan menggunakan algoritma *Naïve Bayes Classifier* dengan model Multinomial. Tujuan dari penelitian ini adalah untuk mengetahui implementasi dari algoritma *Naïve Bayes Classifier*, dan mengetahui hasil perbandingan tingkat akurasi terhadap jumlah dataset yang berbeda-beda. Terdapat 4 tahapan yang akan dilakukan dalam penelitian ini yaitu *Crawling*, *Preprocessing*, *Training-Testing*, dan evaluasi. Dalam tahap evaluasi dilakukan menggunakan confusion matrix. Penggunaan dataset dalam penelitian memiliki 3 dataset yang berbeda jumlahnya, yaitu 1000 dataset, 2000 dataset, dan 3000 dataset.

Hasil pengujian pada penelitian ini menunjukkan bahwa algoritma *Naïve Bayes Classifier* dapat dengan baik diimplementasikan dalam analisis sentimen. Dalam penelitian ini juga menyimpulkan bahwa jumlah dataset yang digunakan sangat berpengaruh terhadap tingkat akurasi dari algoritma *Naïve Bayes Classifier*. Hal ini dibuktikan dengan Hasil pengujian dataset yang berjumlah 3000 data memiliki tingkat akurasi 85%, Sedangkan untuk dataset yang berjumlah lebih sedikit yaitu 1000 data memiliki tingkat akurasi 80%.

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

ABSTRACT

In recent years, many new internet providers have emerged that have attracted many internet users in Indonesia, one of which is by.u. Many of the users give their opinions through social media, especially twitter, from various user opinions there are tweets on twitter containing positive and negative words towards the by.u provider. From this problem, research is conducted using sentiment analysis. To get the best evaluation results, a comparison of training data and test data is carried out. The stages of the research consist of retrieval of data from social media twitter, preprocessing data, weighting TF-IDF data and classification using the Naïve Bayes Classifier algorithm.

In this study, sentiment analysis was carried out on the provider by.u using the Naïve Bayes Classifier algorithm with the Multinomial model. The purpose of this study was to determine the implementation of the Naïve Bayes Classifier algorithm, and to find out the results of the comparison of the level of accuracy against the number of different datasets. There are 4 stages to be carried out in this research, namely Crawling, Preprocessing, Training-Testing, and evaluation. In the evaluation stage, a confusion matrix is used. The use of datasets in this study has 3 different datasets, namely 1000 datasets, 2000 datasets, and 3000 datasets.

The test results in this study indicate that the Naïve Bayes Classifier algorithm can be well implemented in sentiment analysis. This study also concludes that the number of datasets used greatly affects the accuracy of the Naïve Bayes Classifier algorithm. This is evidenced by the results of testing the 3000 datasets that have an accuracy rate of 85%, while for the datasets which amount to less, 1000 data have an 80% accuracy rate.

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