

**IMPLEMENTASI ALGORITMA *NAÏVE BAYES* DAN
ALGORITMA C4.5 UNTUK MELAKUKAN ANALISIS
SENTIMEN TERHADAP ULASAN KOMENTAR
PENGGUNA TIKTOK DI *GOOGLE PLAY STORE***

SKRIPSI

Oleh:

DHEA PUTRI APRILYANA

201910225007



**PROGRAM STUDI INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS BHAYANGKARA JAKARTA RAYA
2023**

LEMBAR PERSETUJUAN PEMBIMBING (DIGUNAKAN UNTUK TUGAS AKHIR)

Judul Tugas Akhir : Implementasi Algoritma *Naive Bayes* dan Algoritma
C4.5 untuk Melakukan Analisis Sentimen terhadap
Ulasan Komentar Pengguna *Tiktok* di *Google*
Play Store

Nama Mahasiswa : DHEA PUTRI APRILYANA

Nomor Pokok Mahasiswa : 201910225052

Program Studi/Fakultas : Informatika/Ilmu Komputer

Tanggal Lulus Ujian Skripsi : 04 Agustus 2023

Jakarta, 04 Agustus 2023

MENYETUJUI

Pembimbing I

Pembimbing II

Wowon Priatna, S.T., M.Ti.
NIDN: 0429118007

Siti Setiawati, S.Pd., M.Pd.
NIDN: 0313107904

LEMBAR PENGESAHAN

Judul Skripsi : IMPLEMENTASI ALGORITMA *NAÏVE BAYES*
DAN ALGORITMA C4.5 UNTUK
MELAKUKAN ANALISIS SENTIMEN
TERHADAP ULASAN KOMENTAR
PENGGUNA TIKTOK DI *GOOGLE PLAY*
STORE

Nama Mahasiswa : DHEA PUTRI APRILYANA

Nomor Induk Mahasiswa : 201910225007

Program Studi/Fakultas : Informatika/Illmu Komputer

Tanggal Lulus Ujian Skripsi : 04 Agustus 2023

Jakarta, 08 Agustus 2023

MENGESAHKAN

Ketua Tim Penguji : Dr. Dra. Tyastuti Sri Lestari, MM.
NIDN: 0327036701

Penguji II : Khairunnisa Fadhillah Ramdhania, S.Si., M.Si.
NIDN: 0328039201

Penguji III : Wowon Priatna, S.T., M.Ti.
NIDN: 0429118007

MENGETAHUI,

Ketua
Program Studi Informatika

Dekan

Fakultas Ilmu Komputer

Ahmad Fathurrozi, S.E., M.M.S.I
NIP. 2012486

Dr. Dra. Tyastuti Sri Lestari, M.M.
NIP. 1408206



UNIVERSITAS BHAYANGKARA JAKARTA RAYA
FAKULTAS ILMU KOMPUTER
PROGRAM STUDI INFORMATIKA

LEMBAR PERNYATAAN BUKAN PLAGIASI

Yang bertanda tangan dibawah ini :

Nama : Dhea Putri Aprilyana
NPM : 201910225007
Program Studi : Informatika
Fakultas : Ilmu Komputer
Judul Tugas Akhir : Implementasi Algoritma *Naïve Bayes* dan Algoritma C4.5
Untuk Melakukan Analisis Sentimen Terhadap Ulasan
Komentar Pengguna TikTok di *Google Play Store*

Dengan ini menyatakan bahwa hasil penulisan skripsi yang telah saya buat ini merupakan **hasil karya saya sendiri dan benar keasliannya**. Apabila dikemudian hari penulisan skripsi ini merupakan plagiat atau penjiplakan terhadap karya orang lain, maka saya bersedia mempertanggungjawabkan sekaligus bersedia menerima sanksi berdasarkan tata tertib di Universitas Bhayangkara Jakarta Raya.

Demikian pernyataan ini saya buat dalam keadaan sadar dan tidak dipaksakan dari pihak manapun.

Jakarta, 10 Agustus 2023
Penulis



Dhea Putri Aprilyana
201910225007

ABSTRAK

Dhea Putri Aprilyana. 201910225007. Implementasi Algoritma *Naïve Bayes* Dan Algoritma C4.5 Untuk Melakukan Analisis Sentimen Terhadap Ulasan Komentar Pengguna TikTok Di *Google Play Store*. Fakultas Ilmu Komputer. Bekasi. Universitas Bhayangkara Jakarta Raya. 2023.

TikTok adalah aplikasi populer dikalangan anak muda. TikTok merupakan aplikasi yang pertama kali diluncurkan di China sebelum hadir di Indonesia pada akhir tahun 2017. Sayangnya, popularitas TikTok tidak hanya berdampak positif tetapi juga berdampak negatif karena kurangnya citra diri individu tersebut, misalnya mengenakan pakaian seksi dan menari dengan gerakan erotis dan tidak pantas. Hal ini didasari oleh banyaknya komentar pengguna TikTok yang bersifat positif dan negatif. Maka diperlukan metode untuk dapat mengklasifikasikan suatu ulasan secara otomatis dengan analisis sentimen. Tujuan dari penelitian ini adalah mengklasifikasikan ulasan komentar pengguna TikTok di *Google Play Store* dengan menggunakan algoritma *Naïve Bayes* dan C4.5. Penelitian ini menggunakan 1330 data yang dimana 602 data negatif dan 728 data positif. Hasil yang menunjukkan dengan algoritma *Naïve Bayes* menghasilkan nilai akurasi 79%, *precision* 79%, *recall* 78%, dan *F1 Score* 78%. Sedangkan algoritma C4.5 menghasilkan nilai akurasi 68%, *precision* 68%, *recall* 68%, dan *F1 Score* 68%. Dapat di simpulkan bahwa algoritma *Naïve Bayes* merupakan algoritma terbaik dibandingkan algoritma C4.5. Algoritma *Naïve Bayes* mendapatkan nilai akurasi sebesar 79%, dimana memiliki nilai yang tinggi dibandingkan *precision*, *recall*, dan *f1-score*. Maka akurasi bisa digunakan sebagai acuan performa algoritma *Naïve Bayes*. Untuk algoritma C4.5 memiliki akurasi 68% yang dimana memiliki nilai sama dengan *precision*, *recall*, dan *f1-score*. Maka akurasi bisa digunakan sebagai acuan performa algoritma C4.5

Kata kunci: TikTok, Implementasi, Analisis Sentimen, *Google Play Store*, *Naive Bayes*, C4.5

ABSTRACT

Dhea Putri Aprilyana. 201910225007. *Implementation of the Naïve Bayes Algorithm and the C4.5 Algorithm to Perform Sentiment Analysis of TikTok User Comment Reviews on the Google Play Store. Faculty of Computer Science. Bekasi. Jakarta Bhayangkara University. 2023.*

TikTok is a popular application among young people. TikTok was an application initially launched in China before landing in Indonesia at the end of 2017. Unfortunately, the popularity of TikTok stems from personal lack of self-image, for example wearing sexy clothes, dancing in erotic and inappropriate moves. This is based on many positive and negative comments from TikTok users. So we need a way to automatically classify reviews through sentiment analysis. The purpose of this study is to classify TikTok user comments on Google Play Store using Naive Bayes and C4.5 algorithms. This study used 1330 data, of which 602 data were negative and 728 data were positive. The results show that the Naive Bayes algorithm produces accuracy values of 79%, 79% precision, 78% recall, and 78% F1 score. The C4.5 algorithm produces 68% accuracy, 68% precision, 68% recall, and 68% F1 score. We can conclude that the Naive Bayes algorithm is the best algorithm compared to the C4.5 algorithm. The Naive Bayes algorithm achieves an accuracy value of 79%. Where it has a higher value compared to precision, recall, and f1-score. Thus, accuracy can be used as a reference for the performance of the Naive Bayes algorithm. For the C4.5 algorithm, it has an accuracy of 68%, which is the same value as precision, recall, and f1-score. Hence, accuracy can be used as a benchmark for the performance of the C4.5 algorithm.

Keywords: TikTok, Implementation, Sentiment Analysis, Google Play Store, Naive Bayes, C4.5

**LEMBAR PERNYATAAN PERSETUJUAN PUBLIKASI
KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIK**

Sebagai sivitas akademik Universitas Bhayangkara Jakarta Raya, saya yang bertanda tangan di bawah ini :

Nama : Dhea Putri Aprilyana
NPM : 201910225007
Program Studi : Informatika
Fakultas : Ilmu Komputer
Jenis Karya : Skripsi

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Bhayangkara Jakarta Raya **Hak Bebas Royalti Non-Eksklusif (*Non-Exclusive Royalty-Free Right*)**, atas karya ilmiah saya yang berjudul :

Implementasi algoritma *Naïve Bayes* dan Algoritma C4.5 Untuk Melakukan Analisis Sentimen Terhadap Ulasan Komentar Pengguna TikTok di *Google play store*

beserta perangkat yang ada (bila diperlukan). Dengan hak bebas royalti non-eksklusif ini, Universitas Bhayangkara Jakarta Raya berhak menyimpan, mengalihmediakan, mengelolanya dalam bentuk pangkalan data (*database*), mendistribusikannya dan mempublikasikannya di Internet atau media lain untuk kepentingan akademis tanpa perlu meminta ijin dari saya selama tetap mencantumkan nama saya sebagai penulis dan sebagai pemilik hak cipta.

Segala bentuk tuntutan hukum yang timbul atas pelanggaran hak cipta dalam karya ilmiah ini menjadi tanggung jawab saya pribadi

Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat di : Jakarta
Pada tanggal : 10 Agustus 2023
Yang Menyatakan



Dhea Putri Aprilyana

KATA PENGANTAR

Puji syukur peneliti panjatkan kehadiran Allah SWT, atas segala rahmat, hidayah, dan karunia-Nya sehingga peneliti dapat menyelesaikan penulisan skripsi ini dengan judul “Implementasi Algoritma *Naive Bayes* dan Algoritma C4.5 untuk Melakukan Analisis Sentimen Terhadap Ulasan Komentar Pengguna TikTok di *Google Play Store*”.

Tugas akhir ini ditulis untuk memenuhi syarat dalam menyelesaikan pendidikan Strata Satu (S1) pada Program Studi Informatika. Peneliti menyadari selama peyusunan tugas akhir ini tidak terlepas dari bantuan, bimbingan dan motivasi dari berbagai pihak. Oleh karena itu, peneliti ingin mengucapkan terima kasih yang sebesar-besarnya kepada:

1. Bapak Inspektur Jenderal Polisi Irjen Pol. (Purn) Dr, Drs. H. Bambang Karsono, S.H.,M.M. selaku Rektor Universitas Bhayangkara Jakarta Raya.
2. Ibu Dr. Dra. Tyastuti Sri Lestari, M.M. selaku Dekan Fakultas Ilmu Komputer Universitas Bhayangkara Jakarta Raya.
3. Bapak Ahmad Fathurrozi, S.E., M.M.S.I. selaku Ketua Program Studi Informatika.
4. Bapak Wowon Priatna, S.T.,M.Ti. selaku dosen pembimbing 1 penyusunan laporan skripsi.
5. Ibu Siti Setiawati, S.Pd.,M.Pd. selaku dosen pembimbing 2 penyusunan laporan skripsi.
6. Ibu Hafizah, S.S.,M.Pd. selaku validator instrument dalam penentuan pelabelan data dalam pembuatan laporan skripsi.

7. Kedua orang tua yang selalu memberi semangat dan dukungan, selalu mendoakan setiap harinya agar diberikan kesehatan dan kemudahan dalam menyelesaikan skripsi.
8. Teman – teman ANGK yang telah bersama-sama dalam pengerjaan skripsi ini dan memberi dukungan satu sama lain.
9. Untuk Arief Priyadi Pratama dan Nur Fadila yang sudah memberi dukungan dan support dalam proses penulisan skripsi ini.
10. Untuk Eka Nur A'ini yang telah bersama – sama membantu dalam proses penyelesaian skripsi ini.
11. Untuk Anisa Dwiyantri, Icha Anggi, dan Mila Febriani yang sudah memberi support penulis selama penyelesaian skripsi ini.
12. Kepada diri sendiri, terima kasih telah berjuang sejauh ini dan tetap kuat melewati berbagai cobaan dan rintangan.

Peneliti menyadari bahwa penelitian ini jauh dari kesempurnaan, untuk itu peneliti mohon maaf apabila terdapat kekurangan dan kesalahan baik dari isi maupun tulisan. Oleh karena itu, peneliti mengharapkan kritik dan saran yang membangun.

Bekasi, 05 Juli 2023

Dhea Putri Aprilyana
201910225007

DAFTAR ISI

LEMBAR PERSETUJUAN PEMBIMBING	i
LEMBAR PENGESAHAN	ii
LEMBAR PERNYATAAN BUKAN PLAGIASI	iii
ABSTRAK	iv
ABSTRACT	v
LEMBAR PERNYATAAN PERSETUJUAN PUBLIKASI KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIK.....	vi
KATA PENGANTAR.....	vii
DAFTAR ISI.....	ix
DAFTAR TABEL	xiii
DAFTAR GAMBAR.....	xv
DAFTAR LAMPIRAN.....	xvii
BAB I PENDAHULUAN.....	1
1.1. Latar Belakang	1
1.2. Identifikasi Masalah	5
1.3. Batasan Masalah.....	5
1.4. Rumusan Masalah	6
1.5. Tujuan Penelitian.....	6

1.6.	Manfaat Penelitian.....	6
1.7.	Sistematika Tugas Akhir	7
BAB II TINJAUAN PUSTAKA.....		9
2.1.	Penelitian Terdahulu.....	9
2.2	<i>Data Scientist</i>	12
2.2.1	<i>Data Mining</i>	12
2.2.2	<i>Text Mining</i>	13
2.3	Analisis Sentimen	14
2.4	Algoritma <i>Naïve Bayes</i>	14
2.4.1	Multinomia NB	16
2.5	Algoritma C4.5.....	17
2.6	KDD (<i>Knowledge Discovery in Database</i>)	19
2.7	Text Preprocessing	19
1.	<i>Case Folding</i>	20
2.	<i>Tokenizing</i>	20
3.	<i>Filtering</i>	20
4.	<i>Stemming</i>	21
2.8	Pembobotan <i>Term Frequency-Inverse Document Frequency</i> (TF-IDF) 21	
2.9	<i>Confusion Matrix</i>	22
2.10	<i>Web Scrapping</i>	24

2.11	<i>Google Play Store</i>	25
2.12	<i>Aplikasi TikTok</i>	25
2.13	<i>Media Sosial</i>	26
BABIII METODOLOGI PENELITIAN		28
3.1.	<i>Kerangka Penelitian</i>	28
3.2	<i>Data Selection</i>	30
3.2.1	<i>Scrapping Data</i>	30
3.2.2	<i>Pelabelan Data</i>	30
3.2.	<i>Preprocessing</i>	31
3.2.1.	<i>Case Folding</i>	31
3.2.3.	<i>Tokenizing</i>	33
3.2.4.	<i>Filtering</i>	34
3.2.5.	<i>Stemming</i>	35
3.3	<i>Transformation</i>	35
3.4	<i>Data Mining</i>	35
3.4.1	<i>Algoritma Naïve Bayes</i>	36
3.5.	<i>Algoritma C4.5</i>	37
3.5	<i>Evaluation</i>	38
3.6	<i>Visualisasi Data</i>	38
BAB IV HASIL DAN PEMBAHASAN		39

4.1.	Hasil.....	39
4.1.1	<i>Data Selection</i>	39
4.2	Pembahasan	76
BAB V PENUTUP.....		77
5.1	Kesimpulan.....	77
5.2.	Saran.....	77
DAFTAR PUSTAKA.....		78
LAMPIRAN.....		83



DAFTAR TABEL

Tabel 2.1 Tabel Penelitian Terdahulu	9
Tabel 2.2 Confusion Matrix	22
Tabel 3.1 Data hasil labelling.....	31
Tabel 3.2 Aturan tahapan case folding.....	33
Tabel 3.3 Aturan tahapan tokenizing	34
Tabel 4.1 Pelabelan Data.....	41
Tabel 4.2 Tahapan Case Folding.....	43
Tabel 4.3 Tahapan Tokenizing.....	45
Tabel 4.4 Tahapan Filtering	47
Tabel 4.5 Tahapan Stemming	50
Tabel 4.6 Dokumen Yang Akan di Hitung	52
Tabel 4.7 Tabel Mencari TF dan IDF	53
Tabel 4.8 Tabel Mencari Bobot	54
Tabel 4.9 Tabel Perhitungan Manual untuk Data Latih.....	57
Tabel 4.10 Tabel Tahapan Manual untuk Penentuan Prediksi.....	58
Tabel 4.11 Tabel Perhitungan Prior Probability	59
Tabel 4.12 Tabel Sampel Data Test	60
Tabel 4. 13 Menghitung banyaknya partisi.....	62
Tabel 4.14 Menghitung jumlah entropy dan gain	64
Tabel 4.15 Tabel training data	65
Tabel 4.16 Nilai Data Training	65
Tabel 4.17 Perhitungan Gain untuk simpul akar.....	67

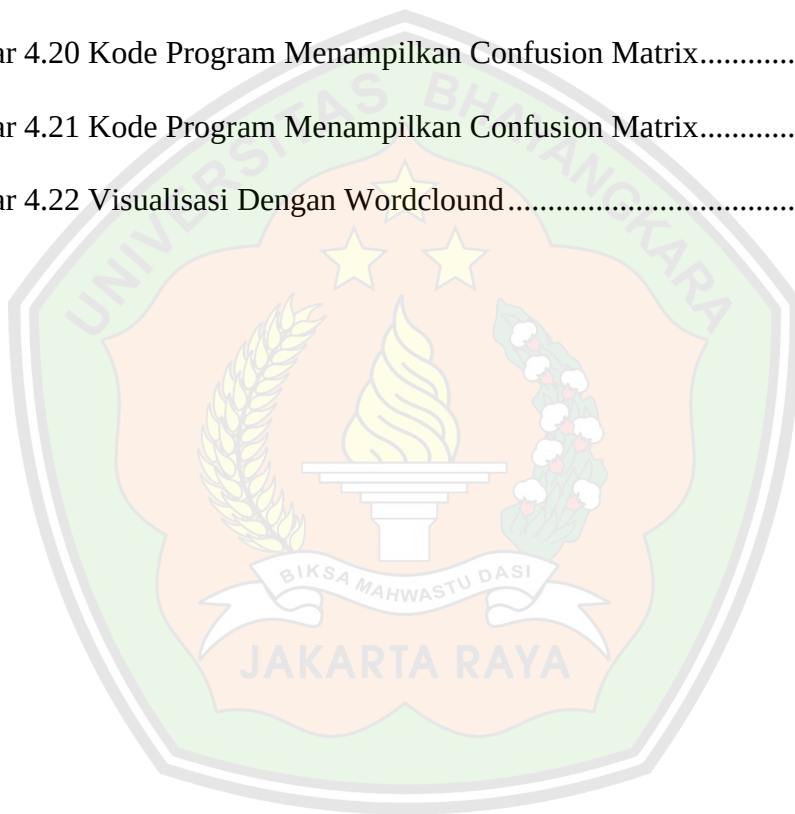
Tabel 4.18 Nilai data training.....	68
Tabel 4.19 Perhitungan gain untuk simpul akar.....	68
Tabel 4.20 Nilai data training.....	69
Tabel 4. 21 Perhitungan Gain simpul Apk.....	69
Tabel 4.22 Tabel confusion matrix naïve bayes.....	71
Tabel 4.23 Tabel Confusion Matrix C4.5	73
Tabel 4. 24 Hasil Confusion Matrix.....	76



DAFTAR GAMBAR

Gambar 1.1 Data Ulasan dan Jumlah Unduhan di Google Play Store.....	2
Gambar 3.1 Kerangka Penelitian	28
Gambar 3.2 Flowchart Case Folding	32
Gambar 3.3 Flowchart Tokenizing	33
Gambar 3.4 Flowchart Filtering	34
Gambar 3.5 Flowchart Stemming	35
Gambar 3.6 Tahapan Naïve Bayes.....	36
Gambar 3.7 Tahapan C4.5	37
Gambar 4.1 Proses Pengambilan Data Komentar Dari Google Play Store.....	40
Gambar 4.2 hasil data ulasan yang sudah di scrapping.....	40
Gambar 4.3 Kode Program Case Folding	44
Gambar 4.4 Hasil Kode Program Dari Case Folding.....	44
Gambar 4.5 Kode Program Tokenizing	46
Gambar 4.6 Hasil Proses Tokenizing.....	47
Gambar 4.7 Kode Program Filtering.....	49
Gambar 4.8 Hasil Proses Filtering	49
Gambar 4.9 Kode Program Stemming	51
Gambar 4.10 Hasil proses filtering	51
Gambar 4.11 Kode Program Tahapan Merubah Menjadi Numpy Array.....	55
Gambar 4.12 Hasil Tahapan Merubah Data Menjadi Numpy Array Komentar ...	55
Gambar 4.13 Hasil Tahapan Merubah Data Menjadi Numpy Arrayy Kolom Sentimen.....	55

Gambar 4.14 Kode program Splitting Dataset.....	56
Gambar 4. 15 Kode Program Modelling Naïve Bayes	56
Gambar 4.16 Kode Program Library C4.5.....	62
Gambar 4.17 Hasil perhitungan nilai gain	68
Gambar 4.18 Hasil perhitungan nilai gain	69
Gambar 4.19 Hasil perhitungan nilai Gain simpul apk.....	70
Gambar 4.20 Kode Program Menampilkan Confusion Matrix.....	72
Gambar 4.21 Kode Program Menampilkan Confusion Matrix.....	75
Gambar 4.22 Visualisasi Dengan Wordcloud.....	75



DAFTAR LAMPIRAN

Lampiran 1	Plagiarisme
Lampiran 2	Biodata Mahasiswa
Lampiran 3	Kartu Bimbingan Skripsi
Lampiran 4	Surat Rekomendasi dari Pembimbing
Lampiran 5	Surat Pertanyaan Validitas Instrumen
Lampiran 6	Lembar Validator Instrumen

