PELITA TEKNOLOGI
Journal homepage: jurnal.pelitabangsa.ac.id, p-ISSN: 2301-475X, e-ISSN: 2656-7059

Implementasi Algoritma *Naïve Bayes* dan Algoritma C4.5 Untuk Melakukan Analisis Sentimen terhadap Ulasan Komentar Pengguna TikTok di *Google Play Store*

Dhea Putri Aprilyana[1], Wowon Priatna[2], Siti Setiawati[3]

Program Studi Informatika, Universitas Bhayangkara Jakarta Raya
Jl. Raya Perjuangan No.81, RT.003/RW.002, Marga Mulya, Kec. Bekasi Utara, Kota Bks, Jawa Barat 17143
Korespondensi email: wowon.priatna@dsn.ubhrajaya.ac.id

Abstrak

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 79.00%, 79.00% precision, 78.00% recall, and 78.00% F1 score. The C4.5 algorithm produces 68.00% accuracy, 68.00% precision, 68.00% recall, and 68.00% 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 79.00%.

Informasi Artikel

Diterima: 16 Juli 2024
Direvisi: 21 Agustus 2024
Dipublikasikan: 21 Maret 2024

Kata Kunci

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

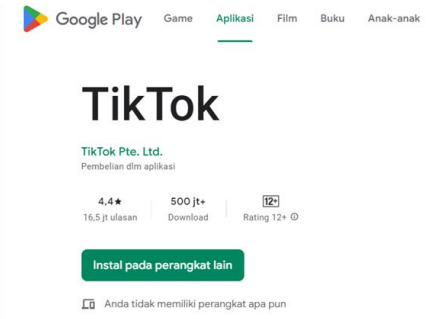
I. Pendahuluan

Teknologi sekarang sudah sangat maju, banyak media yang bisa digunakan untuk melakukan komunikasi, media ini sangat mudah digunakan dengan menghubungkan ke internet [1]. Sangat ini, berkat perkembangan teknologi yang

PELITA TEKNOLOGI
Journal homepage: jurnal.pelitabangsa.ac.id, p-ISSN: 2301-475X, e-ISSN: 2656-7059

sangat pesat, masyarakat sangat mudah untuk berkomunikasi menggunakan salah satu teknologi yang modern yaitu smartphone. Smartphone ini sangat memungkinkan masyarakat untuk mengakses media dengan sangat mudah dan masyarakat dapat menggunakannya untuk berkomunikasi, seperti melalui Facebook, Instagram, Twitter, dll. Media ini memiliki berbagai fungsi. Salah satu aplikasi yang saat ini sedang banyak mendapat perhatian adalah aplikasi TikTok. Aplikasi TikTok ini memungkinkan pengguna membuat video berdurasi pendek dan cepat [2] [3].

TikTok adalah aplikasi populer di kalangan anak muda. TikTok merupakan aplikasi yang pertama kali diluncurkan di China sebelum hadir di Indonesia pada akhir tahun 2017. Aplikasi TikTok ini resmi diluncurkan oleh Zhang Yiminy pada tahun 2016 [4]. Aplikasi ini dapat menambahkan banyak fitur seperti, Menambahkan musik dalam video, mengubah suara, filter, tambahan efek dan stiker, dll. Aplikasi TikTok ini juga mendorong para penggunanya untuk berkreasi saat membuat video. Sementara itu, Aplikasi ini juga sekarang menjadi aplikasi yang populer di kalangan masyarakat, bahkan di masa pandemi jumlah pengguna TikTok meningkat 20% dibandingkan dengan biasanya [5]. Rata-rata usia pengguna TikTok ini berusia hampir sama dengan rata-rata usia anak di bawah umur dan bahkan kebanyakan orang dewasa [2]. 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. Ironisnya, banyak pengguna terutama remaja yang mencoba untuk mengikuti hal ini [6]. Alasan mengambil penelitian ini adalah dikarenakan aplikasi TikTok saat ini banyak digunakan. Di Google Play Store sendiri, aplikasi TikTok sudah diunduh lebih dari 500jt+ dan memiliki 16.5jt rating pengguna. Ulasan data penelitian ini dengan menggunakan komentar dalam bahasa Indonesia.

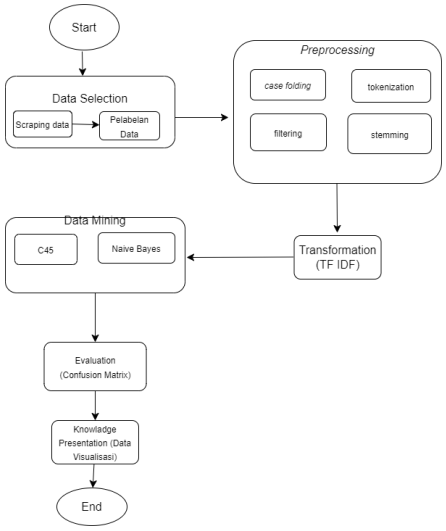


Gambar 1 Data Ulasan dan Jumlah Unduhan di Google Play Store

II. Metode Penelitian

2.1 Tahapan Penelitian

Pada tahap penelitian ini digunakan metode *Knowledge Discovery in Database*. Alur kerja penelitian ini ditunjukkan pada gambar 2.



Gambar 2 Tahapan alur kerangka penelitian

PELITA TEKNOLOGI
Journal homepage: jurnal.pelitabangsa.ac.id, p-ISSN: 2301-475X, e-ISSN: 2656-7059

Berikut uraian alur penelitian yang menerapkan metode KDD:

2.1 Data Selection

Dalam teknik ini di bagi menjadi dua, yaitu *scraping data* dan pelabelan data. *Scraping data* adalah proses pengembilan data atau mengekstaksi data dari berbagai sumber yang terdapat di internet. Sedangkan pelabelan data proses memberikan kategori label data yang tidak memiliki label sebelumnya.

2.2 *Preprocessing*

Pada langkah ini, analisis semantik (kebenaran arti) dan sintaktik (kebenaran susunan) teks. Tujuannya adalah untuk mengubah teks menjadi data berkualitas yang akan diproses lebih lanjut. Pada tahap ini akan dilakukan beberapa langkah diantaranya *case folding, cleaning, tokenizing, filtering,* dan *stemming*.

2.3 *Transformation*

Pada tahap transformasi teks atau pembentukan atribut, dilakukan proses untuk mendapatkan representasi dokumen yang diperlukan. Dalam langkah ini, ekstraksi fitur dilakukan menggunakan metode TF-IDF. Term frequency (tf) merupakan sistem pembobotan yang mengukur frekuensi kemunculan term dalam dokumen. Nilai tf yang lebih besar berarti istilah tersebut sering muncul, sehingga dapat dianggap sebagai istilah umum dan nilainya tidak signifikan. Iverse Document Frequency (idf) merupakan frekuensi berbanding terbalik [7] [8].

$$TF = \frac{jumlah\ term\ di\ dok}{jumlah\ seluruh\ term\ di\ dok} \quad (1)$$

$$IDF = \log_2 \frac{jumlah\ seluruh\ dok}{jumlah\ dok\ pada\ tern} \quad (2)$$

$$TF\ IDF = TF * IDF \quad (3)$$

2.4 Data Mining

Pada tahap ini, dilakukan klasifikasi sentimen pada data ulasan yang telah melalui tahapan transformasi menggunakan algoritma *naïve bayes* dan algoritma C45.

2.5 Algoritma *Naïve Bayes*

Algoritma Naïve Bayes adalah algoritma klasifikasi sederhana yang mana menghitung sekumpulan probabilitas dengan metode menjumlah sekumpulan probabilitas dengan metode menjumlahkan serta mengombinasikan nilai dari dataset yang diberikan. Metode Algoritma Naïve Bayes hendak digunakan pada penelitian ini dalam proses klasifikasi pembahasan pada aplikasi TikTok. Meski klasifikasi Algoritma Naïve Bayes dapat dikatakan klasifikasi simple, namun hasil yang didapat dari klasifikasi Algoritma Naïve Bayes ini kerap sekali menggapai performa yang seragam dengan algoritma klasifikasi yang lain semacam Neural Network Classifier & Decision Tree [9] [10] [11].

$$P(Vj) = (|docs\ j|)/(|contoh|) \quad (4)$$

$$P(xi|Vj) = (nk+1)/(n+|kosakata|) \quad (5)$$

Keterangan:

|dosc j| = jumlah dokumen setiap kategori j

|contoh| = jumlah dokumen dari semua kategori

Nk = jumlah frekuensi kemunculan setiap kata

PELITA TEKNOLOGI
Journal homepage: jurnal.pelitabangsa.ac.id, p-ISSN: 2301-475X, e-ISSN: 2656-7059

N = jumlah frekuensi kemunculan kata dari setiap kategori

|kosakata| = jumlah semua kata dari semua kategori

2.4 Algoritma C4.5

Algoritma C45 adalah bagian dari algoritma klasifikasi untuk machine learning dan data mining. C45 adalah algoritma yang sesuai digunakan buat permasalahan klasifikasi pada machine learning serta data mining [12].

$Entropy$ (S) = $\sum_{i=0}^{n} - pi * \log_2 pi$ (6)

Keterangan:
S = Himpunan Kasus
n = Jumlah Partisi S
pi = Proporsi dari Si terhadap S

Gain(S,A) = *Entropy* (S) $- \sum_{i=0}^{n} \frac{|Si|}{|S|} *$ $Entropy\ (S_i)$ (7)

Keterangan:
S = Himpunan Kasus
A = Atribut
n = Jumlah partisi
|Si| = Jumlah Kasus pada partisi ke-i
|S| = Jumlah Kasus dalam S

Split Info (S,A) = $- \sum_{i=0}^{n} \frac{Si}{S} \log_2 \frac{Si}{S}$ (8)

Informasi potensial atau entropy

Gain Ratio (S,A) = $\frac{Gain\ (S,A)}{Split\ Info\ (S,A)}$ (9)

2.5 Evalution

Pada tahap evalution, suatu tabel yang digunakan untuk melakukan evaluasi terhadap kinerja suatu model klasifikasi.

2.6 Visualisasi

Pada tahap visualisasi ini adalah proses menyajikan informasi atau data secara visual menggunakan grafik, diagram, atau representasi lainnya.

2.7 *Sentiment*

Sentimen atau opini masyarakat semakin bertambah luas dan bebas diungkapkan di berbagai media. Sentimen dapat menjadi potesi besar bagi perusahaan yang ingin mengetahui umpan balik (feedback) dari masyarakat terhadap merek dagang mereka [13] [14].

2.8 Analisis Sentimen

Analisis sentimen merupakan suatu bidang ilmu dari data mining, berguna untuk menganalisis, memproses, dan mengekstraksi data tekstual tentang entitas seperti layanan, produk, individu, organisasi, atau masalah dan topik tertentu [1] [15].

III. Hasil dan Pembahasan

3.1 Pengambilan Data

Pengambilan data yang dilakukan menggunakan bantuan Google Collab dengan menggunakan teknik web scrapping. Pada tahapan pengambilan data ini menggunakan situs web resmi Google Play Store yang memuat berbagai ulasan yang ada di kolom komentar aplikasi Tik Tok. Pada proses pengambilan data dari Google Collab dan menentukan berapa data ulasan komentar yang ingin diambil datanya.

3.2 Pelabelan Data

Pada tahapan ini, data yang sudah di *scrapping* akan dipelabelan manual dengan dibagi menjadi kata positif dan negatif. Pada positif ditandai dengan label 0 dan negatif ditandai dengan label 1.

PELITA TEKNOLOGI
Journal homepage: jurnal.pelitabangsa.ac.id, p-ISSN: 2301-475X, e-ISSN: 2656-7059

Tabel 1 Pelabelan Data

Komentar	Label
Dear Tik tok, kenapa setiap saya apload video! secara otomatis video yang saya apload tiba tiba di kompres, ini sangat mengganggu kami para tik toker, mohon untuk segera di perbaiki	Negatif
Tiktok yg udh di update ada versi simpan vidio, setelah video udh disimpan, ada pemberitahuan vidio tersimpan di album, terus cek album rupanya gada, jadi dimana vidio nya	Negatif
Tolong developer, untuk stories bisa diubah ke postingan.. Padahal tidak sengaja ke tekan stories.. Mau diubah ke postingan malah enggak bisa.	Negatif

3.3 Preprocessing

1. Case Folding

Pada proses *case folding* bertujuan untuk mengubah huruf menjadi huruf kecil dan menghapus karakter lain selain huruf dihilangkan.

Tabel 2 Tahapan Case Folding

Video yang udah ditonton tapi belum sempat dilike kalo dicari lagi gak bisa, padahal gak sengaja ilang gegara iklan. Sebel banget gak bisa liat history video yang udah ditonton padahal penasaran. Ada fitur pencarian video yang sudah ditonton 7 hari terakhir tapi gak guna tetep gak ketemu	video yang udah ditonton tapi belum sempat dilike kalo dicari lagi gak bisa, padahal gak sengaja ilang gegara iklan. sebel banget gak bisa liat history video yang udah ditonton padahal penasaran. ada fitur pencarian video yang sudah ditonton hari terakhir tapi gak guna tetep gak ketemu

2. *Tokenizing*

Tahap *tokenizing* merupakan tahap dimana memisahkan kata dan angka yang penting lainnya dari sebuah teks.

video yang udah ditonton tapi belum sempat dilike kalo dicari lagi gak bisa, padahal gak sengaja ilang gegara iklan. sebel banget gak	video yang udah ditonton tapi belum sempat dilike kalo
bisa liat history video yang udah ditonton padahal penasaran. ada fitur pencarian video yang sudah ditonton hari terakhir tapi gak guna tetep gak ketemu	dicari lagi gak bisa padahal gak sengaja ilang gegara iklan sebel banget gak bisa liat

PELITA TEKNOLOGI
Journal homepage: jurnal.pelitabangsa.ac.id, p-ISSN: 2301-475X, e-ISSN: 2656-7059

	history video yang udah ditonton padahal penasaran ada fitur pencarian video yang sudah ditonton hari terakhir tapi gak guna tetep gak ketemu

3. *Filtering*

Tahap *filtering* adalah tahapan membuang kata-kata yang tidak penting dengan menggunakan *stopwords*.

video yang udah ditonton tapi belum sempat dilike kalo dicari lagi gak bisa padahal gak sengaja ilang gegara iklan sebel	[‘video’, ‘yang’, ‘udah’, ‘ditonton’, ‘tapi’, ‘belum’, ‘sempat’, ‘dilike’, ‘kalo’, ‘dicari’, ‘lagi’, ‘gak’, ‘bisa’, ‘padahal’, ‘gak’, ‘sengaja’, ‘ilang’, ‘gegara’, ‘iklan’, ‘sebel’, ‘banget’, ‘gak’, ‘bisa’, ‘liat’, ‘history’, ‘video’, ‘yang’, ‘udah’, ‘ditonton’, ‘padahal’, ‘penasaran’, ‘ada’, ‘fitur’, ‘pencarian’,
banget gak bisa liat history video yang udah ditonton padahal penasaran ada fitur pencarian video yang sudah ditonton hari terakhir tapi gak guna tetep gak ketemu	‘video’, ‘yang’, ‘sudah’, ‘ditonton’, ‘hari’, ‘terakhir’, ‘tapi’, ‘gak’, ‘guna’, ‘tetep’, ‘gak’, ‘ketemu’]

4. *Stemming*

Tahap *stemming* merupakan tahapan untuk mengubah sebuah kata ke dalam bentuk kata dasar. Dengan menghapus kata imbuhan di depan maupun imbuhan yang ada dibelakang kata.

[‘video’, ‘yang’, ‘udah’, ‘ditonton’, ‘tapi’, ‘belum’, ‘sempat’, ‘dilike’, ‘kalo’, ‘dicari’, ‘lagi’, ‘gak’, ‘bisa’, ‘padahal’, ‘gak’, ‘sengaja’, ‘ilang’, ‘gegara’, ‘iklan’, ‘sebel’, ‘banget’, ‘gak’, ‘bisa’, ‘liat’, ‘history’, ‘video’,	video yang udah tonton tapi belum sempat like kalo cari lagi gak bisa padahal gak sengaja ilang gegara iklan sebel banget gak bisa liat history video yang udah tonton padahal penasaran ada fitur pencarian video yang sudah

PELITA TEKNOLOGI
Journal homepage: jurnal.pelitabangsa.ac.id, p-ISSN: 2301-475X, e-ISSN: 2656-7059

‘yang’, ‘udah’, ‘ditonton’, ‘padahal’, ‘penasaran’, ‘ada’, ‘fitur’, ‘pencarian’, ‘video’, ‘yang’, ‘sudah’, ‘ditonton’, ‘hari’, ‘terakhir’, ‘tapi’, ‘gak’, ‘guna’, ‘tetep’, ‘gak’, ‘ketemu’]	tonton hari terakhir tapi gak guna tetep gak ketemu

3.4 Pombobotan Term Frequenct-Inverse Document Frequency (TF-IDF)

Pada tahapan ini, data yang sudah di *preprocessing* kemudian akan dihitung menggunakan rumus TF-IDF sehingga menghasilkan *vector* yang sudah terbobot. Adapun TF (*Term Frequency*) adalah fruekensi mengacu pada jumlah atau kejadian berulang suatu *term* yang muncul dalam dokumen yang relavan. Sedangkan IDF (*Inverse Document Frequency*) merupakan suatu evaluasi dilakukan untuk menghitung sejumlah mana *term* terbesar secara merata di dalam koleksi dokumen yang relavan. Sebagai tahap awal dalam proses pembobotan kata, dilakukan perhitungan TF (*Term Frequency*).

Tabel 3 Dokumen yang akan dihitung

D1	knp saya instal tiktok gk live tolonng baik
D2	kalo ngetag temen knp ga masuk ya tag an nya
D3	apk nya bagus banget
D4	Istimewa
D5	bagus banget apk nya
D6	Bagus orang yg gabut
D7	apk nya njrr
D8	kecewa tiktok susah fyp bagus
D9	tolong pulih perangkat dan efek belle tiktok saya
D10	kalo fyp nya cewe telanjang ae

Selanjutnya akan dilakukan perhitungan untuk mencari tf. Bisa dilihat pada tabel 4. Setelah mendapatkan hasil *Term Frequency* (TF) dan *Document Frequencty* (DF), tahapan selanjutnya menghitung nilai *weight* (W) menggunakan persamaan 3. Berikut hasil perhitungan manual nilai *weight* (W) pada tabel 5.

Tabel 4 Tabel mencari tf

Token	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	df	D/df	IDF (log D/df)
knp	1	1	0	0	0	0	0	0	0	0	2	5	0,698
saya	1	0	0	0	0	0	0	0	1	0	2	5	0,698
instal	1	0	0	0	0	0	0	0	0	0	1	10	1
tiktok	1	0	0	0	0	0	0	1	1	0	3	3,33	0,522

PELITA TEKNOLOGI
Journal homepage: jurnal.pelitabangsa.ac.id, p-ISSN: 2301-475X, e-ISSN: 2656-7059

Token	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	df	D/df	IDF (log D/df)
gk	1	0	0	0	0	0	0	0	0	0	1	10	1
live	1	0	0	0	0	0	0	0	0	0	1	10	1
tolong	1	0	0	0	0	0	0	0	1	0	2	5	0,698
baik	1	0	0	0	0	0	0	0	0	0	1	10	1
kalo	0	1	0	0	0	0	0	0	0	1	2	5	0,698
ngetag	0	1	0	0	0	0	0	0	0	0	1	10	1
temen	0	1	0	0	0	0	0	0	0	0	1	10	1
ga	0	1	0	0	0	0	0	0	0	0	1	10	1
masuk	0	1	0	0	0	0	0	0	0	0	1	10	1
ya	0	1	0	0	0	0	0	0	0	0	1	10	1
tag	0	1	0	0	0	0	0	0	0	0	1	10	1
an	0	1	0	0	0	0	0	0	0	0	1	10	1
nya	0	0	1	0	1	0	1	0	0	1	4	2,5	0,397
apk	0	0	1	0	1	0	1	0	0	0	3	3,33	0,522
bagus	0	0	1	0	1	1	0	1	0	0	4	2,5	0,397
banget	0	0	1	0	1	0	0	0	0	0	2	5	0,698
istimewa	0	0	0	1	0	0	0	0	0	0	1	10	1
njrr	0	0	0	0	0	0	1	0	0	0	1	10	1
kecewa	0	0	0	0	0	0	0	1	0	0	1	10	1
susah	0	0	0	0	0	0	0	1	0	0	1	10	1
fyp	0	0	0	0	0	0	0	1	0	1	2	5	0,698
pulih	0	0	0	0	0	0	0	0	1	0	1	10	1
perangkat	0	0	0	0	0	0	0	0	1	0	1	10	1
dan	0	0	0	0	0	0	0	0	1	0	1	10	1
efek	0	0	0	0	0	0	0	0	1	0	1	10	1
belle	0	0	0	0	0	0	0	0	1	0	1	10	1
cewe	0	0	0	0	0	0	0	0	0	1	1	10	1
telanjang	0	0	0	0	0	0	0	0	1	0	1	10	1
ae	0	0	0	0	0	0	0	0	1	0	1	10	1
orang	0	0	0	0	0	1	0	0	0	0	1	10	1
yg	0	0	0	0	0	1	0	0	0	0	1	10	1
gabut	0	0	0	0	0	1	0	0	0	0	1	10	1

Tabel 5 Tabel mencari IDF

Token	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10
knp	0,698	0,698	0	0	0	0	0	0	0	0
saya	0,698	0	0	0	0	0	0	0	0,698	0
instal	1	0	0	0	0	0	0	0	0	0
tiktok	0,522	0	0	0	0	0	0	0,522	0,522	0
gk	1	0	0	0	0	0	0	0	0	0
live	1	0	0	0	0	0	0	0	0	0
tolong	0,698	0	0	0	0	0	0	0	0,698	0
baik	1	0	0	0	0	0	0	0	0	0
kalo	0	0,698	0	0	0	0	0	0	0	0,698

PELITA TEKNOLOGI
Journal homepage: jurnal.pelitabangsa.ac.id, p-ISSN: 2301-475X, e-ISSN: 2656-7059

Token	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10
ngetag	0	1	0	0	0	0	0	0	0	0
temen	0	1	0	0	0	0	0	0	0	0
Ga	0	1	0	0	0	0	0	0	0	0
masuk	0	1	0	0	0	0	0	0	0	0
ya	0	1	0	0	0	0	0	0	0	0
tag	0	1	0	0	0	0	0	0	0	0
an	0	1	0	0	0	0	0	0	0	0
nya	0	0	0,397	0	0,397	0	0,39794	0	0	0,397
apk	0	0	0,522	0	0,522	0	0,522879	0	0	0
bagus	0	0	0,397	0	0,397	0,397	0	0,397	0	0
banget	0	0	0,698	0	0,698	0	0	0	0	0
istimewa	0	0	0	1	0	0	0	0	0	0
njrr	0	0	0	0	0	0	1	0	0	0
kecewa	0	0	0	0	0	0	0	1	0	0
susah	0	0	0	0	0	0	0	1	0	0
fyp	0	0	0	0	0	0	0	0,698	0	0,698
pulih	0	0	0	0	0	0	0	0	1	0
perangkat	0	0	0	0	0	0	0	0	1	0
dan	0	0	0	0	0	0	0	0	1	0
efek	0	0	0	0	0	0	0	0	1	0
belle	0	0	0	0	0	0	0	0	1	0
cewe	0	0	0	0	0	0	0	0	0	1
telanjang	0	0	0	0	0	0	0	0	1	0
ae	0	0	0	0	0	0	0	0	1	0
orang	0	0	0	0	0	1	0	0	0	0
yg	0	0	0	0	0	1	0	0	0	0
gabut	0	0	0	0	0	1	0	0	0	0

3.5 Implementasi Algoritma *Naïve Bayes*

Pada proses *training model naïve bayes*, proses tersebut dilakukan dengan menggunakan bantuan *library* pada bahasa pemograman *Python* untuk proses *training model* tersebut. Adapun *library scikit-learn* yang digunakan di antarannya adalah *Naïve Bayes, MultinomiaNB.*

Tabel 6 Tabel perhitungan manual untuk data latih

D	Komentar	Kategori
D1	knp saya instal tiktok gk live tolong baik	Positif
D2	kalo ngetag temen knp ga masuk ya tag an nya	Negatif
D3	apk nya bagus banget	Positif
D4	istimewa	Positif
D5	bagus banget apk nya	Positif
D6	bagus orang yg gabut	Positif
D7	apk nya njrr	Negatif
D8	kecewa tiktok susah fyp bagus	Negatif

Sampel data *train* yang nantinya akan melalui tahap penentuan prediksi dan actual untuk mengetahui kata yang ada

PELITA TEKNOLOGI
Journal homepage: jurnal.pelitabangsa.ac.id, p-ISSN: 2301-475X, e-ISSN: 2656-7059

pada setiap kalimat dapat masuk ke setiap class. Tahapan manual penentuan prediksi dan actual dapat dilihat pada tabel 7

Tabel 7 Tabel tahapan manual untuk penentuan prediksi

Prediksi	Aktual		Total
	Positif	Negatif	
knp	1	0	1
saya	1	0	1
instal	1	0	1
tiktok	1	1	2
gk	1	0	1
live	1	0	1
tolong	1	0	1
baik	1	0	1
kalo	0	1	1
ngetag	0	1	1
temen	0	1	1
ga	0	1	1
masuk	0	1	1
ya	0	1	1
tag	0	1	1
an	0	1	1
nya	2	2	4
apk	2	1	3
bagus	2	0	2
banget	2	0	2
istimewa	1	0	1
orang	1	0	1
yg	1	0	1
gabut	1	0	1
njrr	0	1	1
kecewa	0	1	1
susah	0	1	1
fyp	0	1	1
Jumlah Term	15	15	35

Pada tabel diperoleh jumlah *term* positif sebanyak 15 *term* dan jumlah *term* negatif sebanyak 15 *term* dengan total kata 35 kata. Setelah menentukan prediksi dan aktual pada setiap kalimat, selanjutnya adalah menghitung *prior probability*. Adapun tahapan perhitungan tahapan *prior probability* bisa dilihat pada tabel 8.

Tabel 8 Tabel perhitungan prior probability

	$P_{(Positif)}$	$P_{(Negatif)}$
Prior Probability	$\frac{15}{35} = 0{,}42857142$	$\frac{15}{35} = 0{,}42857142$

3.6 Implementasi Algoritma C4.5

Pada proses *training model C45*, proses tersebut dilakukan dengan menggunakan bantuan *library* pada bahasa pemograman *Python* untuk proses *training model* tersebut. Adapun *library scikit-learn* yang digunakan di antarannya adalah *Decision Tree*, *Decision Tree Classifier*. Pada tabel 9 merupakan perhitungan manual dari algoritma C4.5.

Tabel 9 Menghitung banyaknya partisi

Atribut	Partisi	S	No	Yes
Positif	knp	1	0	1
	saya	1	0	1
	instal	1	0	1
	tiktok	2	1	1
	gk	1	0	1
	live	1	0	1
	tolong	1	0	1
	nya	4	2	2
	apk	3	1	2
	bagus	2	0	2
	banget	2	0	2
	istimewa	1	0	1
	orang	1	0	1
	yg	1	0	1
	gabut	1	0	1
	kalo	1	1	0
	ngetag	1	1	0
	temen	1	1	0
	ga	1	1	0
Negatif	masuk	1	1	0
	ya	1	1	0
	tag	1	1	0
	an	1	1	0

PELITA TEKNOLOGI
Journal homepage: jurnal.pelitabangsa.ac.id, p-ISSN: 2301-475X, e-ISSN: 2656-7059

Atribut	Partisi	S	No	Yes
	njrr	1	1	0
	kecewa	1	1	0
	susah	1	1	0
	fyp	1	1	0
Atribut	Partisi	S	No	Yes
	Jumlah	35	15	19

Pada tabel 10 diperoleh hasil perhitungan *entropy* dan *gain* seperti tabel dibawah ini.

Tabel 10 Menghitung jumlah entropy dan gain

Atribut	Partisi	Kasus (S)	No (S1)	Yes (S2)
Total		35	15	19
Positif	knp	1	0	1
	saya	1	0	1
	instal	1	0	1
	tiktok	2	1	1
	gk	1	0	1
	live	1	0	1
	tolong	1	0	1
	nya	4	2	2
	apk	3	1	2
	bagus	2	0	2
	banget	2	0	2
	istimewa	1	0	1
	orang	1	0	1
	yg	1	0	1
	gabut	1	0	1
	kalo	1	1	0
	ngetag	1	1	0
	temen	1	1	0
	ga	1	1	0
Negatif	masuk	1	1	0
	ya	1	1	0
	tag	1	1	0
	an	1	1	0
	njrr	1	1	0
	kecewa	1	1	0
	susah	1	1	0
	fyp	1	1	0

Pada dokumen yang ada, beberapa kata digunakan sebagai atribut untuk perhitungan *entropy* dan *gain*. Pada tabel 4.11 adalah proses pembentukan pohon keputusan dengan algoritma C4.5 menggunakan data *training* kelas negatif dan positif.

Tabel 11 Tabel training data

No	Data Training	Kelas
1	Apk bagus banget	Positif
2	Bagguss	Positif
3	Tampilan menjadi besar semua	Negatif
4	Bagus banget apk nya!	Positif
5	Kadang malas juga lihat aplikasi tiktok ini, dikit melanggar komunitas padahal Cuma makai musik nya doang	Negatif

Kemudian setiap sampel dimasukkan kedalam array dengan semua atributnya, nilai atributnya adalah biner (0-1), dimana nilai 0 berarti kata tidak muncul dan nilai 1 berarti kata muncul. Pada tabel 12 adalah tabel data *training* dan proses perhitungan gain untuk semua atribut.

Tabel 12 Nilai data training

D	Apk	Bagus	Tampilan	Melanggar	Jumlah	Kelas
D1	1	1	0	0	2	Positif

PELITA TEKNOLOGI
Journal homepage: jurnal.pelitabangsa.ac.id, p-ISSN: 2301-475X, e-ISSN: 2656-7059

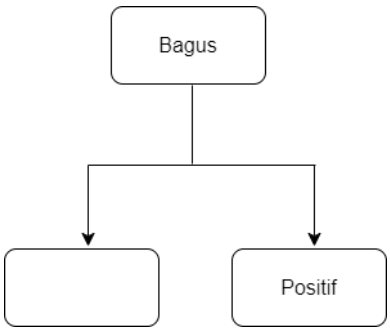
D	Apk	Bagus	Tampilan	Melanggar	Jumlah	Kelas
D2	0	1	0	0	1	Positif
D3	0	0	1	0	1	Negatif
D4	1	1	0	0	2	Positif
D5	0	0	0	1	1	Negatif
Entropy Total	2,451					

Proses perhitungan akan dilakukan sampai dengan ditentukan simpul daun. Selanjutnya pada tabel 13 seluruh hasil perhitungan nilai gain untuk membentuk simpul akar pohon keputusan.

Tabel 13 Perhitungan gain untuk simpul akar

Atribut	Neg (0)	Post (0)	Entropy (0)	Neg (1)	Post (1)	Entropy (1)	Gain
Apk	2	1	0,993	0	2	0,529	2,07
Tampilan	2	0	0,529	0	3	0,529	2,50
Melanggar	0	3	0,907	1	0	0,464	1,82
Gain Tertinggi	2,50						
Node	Bagus						

Dari hasil perhitungan teersebut, nilai gain untuk simpul akar didapatkan *rule tree* seperti pada gambar 3 .



Gambar 3 Hasil perhitungan nilai gain

Karena nilai *entropy* untuk cabang simpul “bagus” dari nilai 0 masih belum sama dengan 0, maka dihitung ulang untuk menentukan node selanjutnya.

Tabel 14 Nilai data training

No	Data training	Kelas
1	Apk bagus banget	Positif
3	Tampilan menjadi besar semua	Negatif
5	Kadang malas juga lihat aplikasi tiktok ini, dikit melanggar komunitas padahal Cuma makai musik nya doang	Negatif

Perhitungan *entropy* total:

Entropy total = 1,585

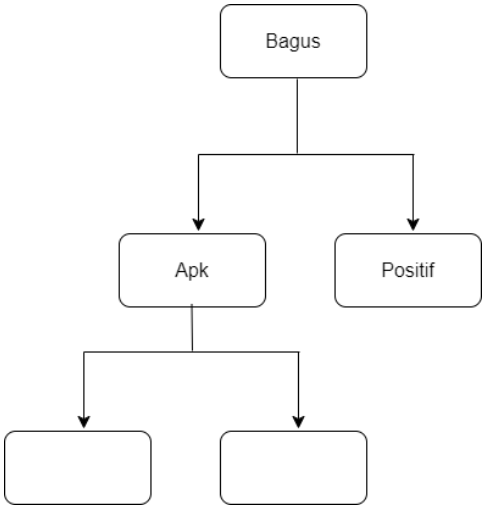
Tabel 15 Perhitungan gain untuk simpul akar

Atribut	Neg (0)	Post (0)	Entropy (0)	Neg (1)	Post (1)	Entropy (1)	Gain
Apk	2	0	0	0	1	0	1,39
Tampilan	1	1	1,06	1	0	0,53	0,70

PELITA TEKNOLOGI
Journal homepage: jurnal.pelitabangsa.ac.id, p-ISSN: 2301-475X, e-ISSN: 2656-7059

Atribut	Neg (0)	Post (0)	Entropy (0)	Neg (1)	Post (1)	Entropy (1)	Gain
Melanggar	1	1	1,06	1	0	0,53	1,06
Gain Tertinggi	1,39						
Node	Apk						

Dari hasil perhitungan di atas, nilai gain untuk menentukan simpul apk didapatkan *rule tree* seperti pada gambar 4.



Gambar 4 Hasil perhitungan nilai gain

Karena nilai *entropy* untuk cabang node "apk" dengan nilai 1 masih belum sama dengan 0, maka akan dihitung ulang untuk menentukan node berikutnya.

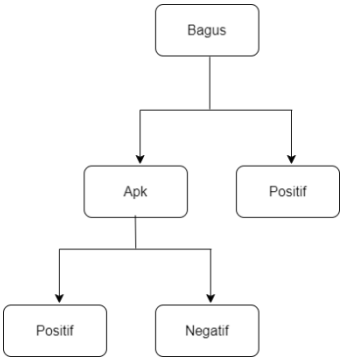
Tabel 16 Nilai data training

D	Tampilan	Melanggar	Jumlah	Kelas
D2	1	0	1	Negatif
D3	0	1	1	Negatif
Entropy Total	0			

Tabel 17 Perhitungan gain simpul apk

Atribut	Neg (0)	Post (0)	Entropy (0)	Neg (1)	Post (1)	Entropy (1)	Gain
Tampilan	1	0	1	1	0	1	0
Melanggar	1	0	1	1	0	1	0

PELITA TEKNOLOGI
Journal homepage: jurnal.pelitabangsa.ac.id, p-ISSN: 2301-475X, e-ISSN: 2656-7059



Gambar 5 Hasil perhitungan gain simpul apk

3.8 Evaluasi model Algoritma *Naïve Bayes*

Evaluasi model dilakukan setelah *training model* selesai dilakukan. Evaluasi model dilakukan untuk mnenghitung performa metode yang digunakan. Kelas sebenernya adalah kelas yang sudah ditentukan nilai TP, FN, FP, TN. Akurasi, *precison*, *recall*, dan f-measure dengan menghitung menggunakan rumus dari masing-masing kelas tersebut. Adapun perhitungan manual dapat dilihat seperti dibawah ini

$$\text{Akurasi} = \frac{TP+TN}{TP+TN+FP+FN} * 100\% = \frac{95+114}{95+114+27+30} * 100\% = \frac{209}{266} * 100\% = 0.79$$

$$\textit{Precision} = \frac{TP}{TP+FP} = \frac{95}{95+27} = 0.78$$

$$\textit{Recall} = \frac{TP}{TP+FN} = \frac{95}{95+30} = 0.76$$

$$\text{f – measure} = 2 * \frac{precision * recall}{preicion+recall} = 2 * \frac{0.78*0.76}{0.78+0.76} = 2 * \frac{0.60}{1.54} = 0.78$$

Jadi pengujian akurasi yang diperoleh dengan menggunakan 1330 data yang terdiri 1064 data latih dan 266 data uji menghasilkan akurasi 79%, *precision* 78%, *recall* 76%, dan f-meaure 0.78.

3.9 Evaluasi model Algoritma C4.5

Evaluasi model dilakukan setelah *training model* selesai dilakukan. Evaluasi model dilakukan untuk mnenghitung performa metode yang digunakan. Kelas sebenernya adalah kelas yang sudah ditentukan nilai TP, FN, FP, TN. Akurasi, *precison*, *recall*, dan f-measure dengan menghitung menggunakan rumus dari masing-masing kelas tersebut.

$$\text{Akurasi} = \frac{TP+TN}{TP+TN+FP+FN} * 100\% = \frac{82+100}{82+100+41+43} * 100\% = \frac{182}{266} * 100\% = 0.69$$

$$\textit{Precision} = \frac{TP}{TP+FP} = \frac{82}{82+41} = 0.68$$

$$\textit{Recall} = \frac{TP}{TP+FN}$$

PELITA TEKNOLOGI
Journal homepage: jurnal.pelitabangsa.ac.id, p-ISSN: 2301-475X, e-ISSN: 2656-7059

$$= \frac{82}{82+43}$$

$$= 0.66$$

f – measure $$=2* \frac{precision*recall}{preicion+recall}$$

$$= 2* \frac{0.67*0.656}{0.67+0.656}$$

$$= 2* \frac{0.43952}{1.326}$$

$$= 0.67$$

Jadi pengujian akurasi dari algoritma C45 yang diperoleh dengan menggunakan 1330 data yang terdiri 1064 data latih dan 266 data uji menghasilkan akurasi 68%, *precision* 67%, *recall* 66 %, dan f-meaasure 67%.

4.0 Visualisasi Data

Pada tahapan ini, penulis ingin mengetahui frekuensi kata yang sering muncul dalam ulasan komentar tiktok di *Google Play Store*. Untuk itu penulis memvisualisasikannya dalam bentuk *wordclound* pada gambar .

Gambar 4 Visualisasi dengan worldcloud

Berdasarkan hasil visualisasi kata dengan *worldclound*, terlihat bahwa kata yang paling banyak dibicarakan dicetak lebih besar dari kata lainnya. Dengan begitu kata video dan tiktok merupakan kata yang sering dibicarakan pada ulasan komentar yang diambil dari aplikasi TikTok di *Google Play Store*.

IV. Kesimpulan

Berdasarkan analisis sentimen yang telah dilakukan peneliti, maka dapat menarik kesimpulan bahwa penelitian ini dilakukan dengan membandingkan dua metode *data mining* yaitu algoritma *naïve bayes* dan algoritma C4.5 dengan menggunakan data ulasan komentar TikTok di *Google Play Store* dengan total data hingga 1330 data ulasan komentar. Data dari hasil klasifikasi *Naïve* Bayes dan C4.5 menggunakan menggunakan model *confusion* matrix yang menghasilkan nilai akurasi, *precision, recall,* dan F1-score dari setiap algoritma. Hasil pengukuran akurasi dengan algoritma *naïve bayes* memberikan hasil akurasi sebesar 79.00% sedangkan algoritma C4.5 sebesar 68.00%. Hal ini menunjukkan bahwa algoritma *naïve bayes* memberikan akurasi yang cukup baik dalam ulasan komentar TikTok di *Google Play Store.*

References

[1] A. I. H. D. P. Valentino Kevin Sitanayah Que, “Analisis Sentimen Transportasi Online Menggunakan Support Vector Machine Berbasis Particle Swarm Optimazion (Online Transportation Sentiment Analysis Using Support Vector Machine Based on Particle Swarm Optimaztion),” *Jurnal Nasional Teknik Elektro dan Teknologi Informasi*, vol. 9, 2020.

[2] I. P. P. A. Junda Alfiah Zulqornain, “Analisis Sentimen Tanggapan Masyarakat Aplikasi Tiktok Menggunakan,” *Jurnal*

PELITA TEKNOLOGI
Journal homepage: jurnal.pelitabangsa.ac.id, p-ISSN: 2301-475X, e-ISSN: 2656-7059

Pengembangan Teknologi Informasi dan Ilmu Komputer, vol. 5, pp. 2886-2890, 2021.

[3] N. K. W. R. Agus Darmawan, "Implementasi Data Mining Menggunakan Model Svm Untuk Prediksi Kepuasan Pengunjung Taman Tabebuya," *Jurnal String* , vol. 2, pp. 299-307, 2018.

[4] Y. Harfian, "Klasifikasi Sentimen Aplikasi Dompet Digital Dana Pada Komentar Di Instagram Menggunakan Naive Bayes Classifier," Pekanbaru, 2021.

[5] A.Rafiq, "Dampak Media Sosial Terhadap Perubahan Sosial Suatu Masyarakat," vol. 1, pp. 18-29, 2020.

[6] A. R. d. Nurnazmi, "Dampak Aplikasi Tiktok Dalam Proses Sosial Di Kalangan Remaja Kelurahan Rabadompu Timur Kecamatan Raba Kota Bima," *Jurnal Pendidikan Sosiologi*, vol. 4, 2021.

[7] S. S. Sola Fide, "Analisis Sentimen Ulasan Aplikasi Tiktok Di Google Play Menggunakan Metode Support Vector Machine (SVM) Dan Asosiasi," *JURNAL GAUSSIAN*, vol. 10, pp. 346 - 358, 2021.

[8] F. W. E. A. S. ArmyliaMalimbe, "Dampak Penggunaan Aplikasi Online Tiktok (Douyin) Terhadap Minat Belajar di Kalangan Mahasiswa Sosiologi Fakultas Ilmu Sosial Dan Politik Universitas Sam Ratulangi Manado," *JURNAL ILMIAH SOCIETY*, vol. 1, pp. 1-10, 2021.

[9] I. R. I. M. Bintang Zulfikar Ramadhan, "Analisis Sentimen Ulasan Pada Aplikasi E-Commerce Dengan Menggunakan Algoritma Naïve Bayes," *Journal of Applied Informatics and Computing (JAIC)*, vol. 6, 2022.

[10] H. P. D. A. F. N. K. A. Y. P. Y. D. Y. S. Rahmadya Trias Handayanto, "Analisis Sentimen Pada Situs Google Review dengan Naïve Bayes dan Support Vector Machine," *Jurnal Komtika (Komputasi dan Informatika)*, vol. 5, pp. 153-163, 2021.

[11] D. E. R. L. M. Bening Herwijayanti, "Klasifikasi Berita Online dengan menggunakan Pembobotan TF-IDFdan Cosine Similarity," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 2, pp. 306-312, 2018.

[12] E. Fitriani, "Perbandingan Algoritma C4.5 Dan Naïve Bayes Untuk Menentukan Kelayakan Penerima Bantuan Program Keluarga Harapan," *Jurnal Sistem Informasi*, vol. 9, pp. 103-115, 2020.

[13] F. Apif Supriadi, "Implementasi Metode Klasifikasi Naive Bayes Pada Sistem Analisis Opini Pengguna Twitter Berbasis Web," *JURNAL SISTEM INFORMASI STMIK ANTAR BANGSA*, vol. 10, pp. 46-54, 2021.

[14] H. S. P. E. E. P. Billy Gunawan, "Sistem Analisis Sentimen pada Ulasan Produk Mwnggunakan Metode Naive Bayes," *Jurnal Edukasi dan Penelitian Informatika*, vol. 4, 2018.

PELITA TEKNOLOGI
Journal homepage: jurnal.pelitabangsa.ac.id, p-ISSN: 2301-475X, e-ISSN: 2656-7059

[15] S. S. Rakhmi Khalida, "Analisis Sentimen Sistem E-Tilang Menggunakan Algoritma Naive Bayes Dengan Optimalisasi Information Gain," *Journal of Information and Information Security*, vol. 1, pp. 19-26, 2020.