

# Giwa

*by* 1 1

---

**Submission date:** 29-Feb-2024 05:34AM (UTC+0200)

**Submission ID:** 2307559262

**File name:** Giwa\_Salsabila\_Jurnal.pdf (806.67K)

**Word count:** 7206

**Character count:** 40815



Klik di sini dan tuliskan Kategori Artikel

## Perbandingan Kinerja Metode *Random Forest* dan *Naïve Bayes* untuk Analisis Sentimen Masyarakat Terhadap *Childfree* Pada *Twitter*

Giwa Salsabila Ferari Irianto <sup>a,\*</sup>, Fajar Astuti Hermawati <sup>b</sup>

<sup>a,b</sup> Jurusan Teknik Informatika, Fakultas Teknik, Universitas 17 Agustus 1945, Surabaya, Indonesia

### INFORMASI ARTIKEL

#### Sejarah Artikel:

Diterima Redaksi: 00 Februari 0000  
Revisi Akhir: 00 Maret 0000  
Diterbitkan Online: 00 April 0000

### KATA KUNCI

Analisis Sentimen, *Childfree*, *Naïve Bayes*,  
*Random Forest*

### KORESPONDENSI

E-mail: [fajarastuti@untag-sby.ac.id](mailto:fajarastuti@untag-sby.ac.id)\*

### ABSTRACT

Fenomena *childfree* mulai menjadi suatu konsep pernikahan bagi masyarakat ketika suami dan istri memilih dan berkomitmen untuk tidak memiliki anak dalam pernikahannya. Faktor munculnya fenomena tersebut yaitu faktor ekonomi, ketidaksiapan mental dalam mengasuh dan membesarkan anak, tingkat kesibukan pasangan suami-istri dan bahkan isu permasalahan lingkungan sosial. Fenomena *childfree* masih menuai pro dan kontra karena adanya perbedaan pendapat dan sudut pandang mengenai kehadiran anak di dalam pernikahan. Pro dan kontra terhadap fenomena tersebut dapat dilihat pada media sosial *Twitter*. Fokus penelitian ini adalah pada analisis sentimen terhadap pandangan masyarakat mengenai *childfree* di *Twitter*, menggunakan metode analisis *Random Forest* dan *Naïve Bayes*. Dalam mendapatkan data sentimen masyarakat terhadap *childfree* di media sosial *Twitter* pada penelitian ini dilakukan menggunakan teknik *scrapping* berupa *tweet* dari masyarakat Indonesia mengenai *childfree*, selain itu *tweet* tersebut mengandung Bahasa Indonesia. Berdasarkan pengumpulan data tersebut diperoleh sebanyak 1309 *tweet*. Hasil evaluasi menunjukkan bahwa *Random Forest* memiliki tingkat akurasi yang lebih tinggi (81.85%) dibandingkan dengan *Naïve Bayes* (78.75%) ketika *Pos Tagger* digunakan. Namun, tanpa *Pos Tagger*, *Naïve Bayes* mencapai akurasi lebih tinggi (69.72%) daripada *Random Forest* (67.33%). Studi ini memberikan wawasan tentang respons masyarakat terhadap fenomena *childfree* dan perbandingan kinerja dua metode analisis sentimen.

## 1. PENDAHULUAN

Manusia sebagai makhluk sosial memiliki kebutuhan untuk berinteraksi dengan orang lain, melakukan pernikahan dan membangun sebuah keluarga. Namun seiring perkembangan zaman, ilmu pengetahuan dan budaya tujuan membangun keluarga mulai berubah. Salah satu fenomena yang hadir dalam membangun keluarga yaitu fenomena *childfree* [1].

Fenomena *childfree* mulai menjadi suatu konsep pernikahan bagi masyarakat ketika suami dan istri memilih dan berkomitmen untuk tidak memiliki anak dalam pernikahannya[2]. Faktor munculnya fenomena tersebut yaitu faktor ekonomi, ketidaksiapan mental dalam mengasuh dan membesarkan anak, tingkat kesibukan pasangan suami-istri dan bahkan isu permasalahan lingkungan sosial [3]. Fenomena *childfree* masih menuai pro dan kontra karena adanya perbedaan pendapat dan sudut pandang mengenai kehadiran anak di dalam pernikahan[1]. Pro dan kontra terhadap fenomena tersebut dapat dilihat pada media sosial *Twitter*. *Twitter* merupakan salah satu media sosial



**2.2. Preprocessing**

Selanjutnya, dalam tahap berikutnya dilakukan proses Preprocessing, yang merupakan langkah penting dalam membersihkan dan mengelola data sebelum dilakukan pembobotan dan analisis[13]. Pada tahap preprocessing ini, digunakan beberapa metode seperti casefolding, tokenize, penghapusan stopwords, stemming, serta proses dengan dan tanpa penggunaan PoS Tagging. Selain itu, dilakukan ekstraksi data untuk memastikan keakuratan.

**2.3. Pelabelan**

Langkah berikutnya adalah melakukan pelabelan data untuk memisahkan sentimen. Proses pelabelan ini menggunakan Corpus, yang merupakan kumpulan data teks yang digunakan untuk mencocokkan dan menentukan kelas labelnya[14]. Corpus ini berperan penting dalam mengidentifikasi dan memetakan sentimen dari setiap data teks. Melalui proses ini, data akan diberikan label yang mencerminkan sentimen positif, negatif, atau netral [15].

**2.4. Split Data**

Setelah proses pelabelan selesai, langkah selanjutnya adalah melakukan pembagian data menjadi dua bagian, yaitu data training dan data testing. Data training akan digunakan untuk melatih model dalam melakukan analisis sentimen, sementara data testing akan berfungsi untuk menguji model dan mengevaluasi performanya. Jumlah sampel data yang akan digunakan dalam tahap ini sebanyak 6, di mana masing-masing kelas akan diwakili oleh 2 data. Dataset ini akan dibagi menjadi dua bagian, yakni bagian training dan bagian testing, guna memastikan keberagaman representasi kelas dalam kedua set data tersebut. Dengan pembagian yang seimbang, diharapkan model dapat belajar dengan baik dan menghasilkan hasil analisis sentimen yang akurat pada data baru yang belum pernah dilihat sebelumnya.

**2.5. Pembobotan Data TF-IDF**

Setelah tahap pelabelan data dan pembagian menjadi data training dan data testing, langkah selanjutnya adalah melakukan pembobotan tanpa menggunakan PoS Tagging. Pembobotan ini bertujuan untuk mengubah data teks menjadi representasi numerik, mempermudah perhitungan pada tahap berikutnya menggunakan metode TF-IDF. TF-IDF digunakan untuk menilai seberapa penting suatu kata dalam suatu dokumen dengan memadukan konsep frekuensi kata (Term Frequency, TF) dan invers frekuensi dokumen (Inverse Document Frequency, IDF) [16]. TF mengukur seberapa sering suatu kata muncul dalam dokumen tertentu, sementara IDF mengukur seberapa unik atau penting kata tersebut dalam seluruh korpus. Berikut merupakan penjabaran dari proses pembobotan TF-IDF menggunakan PoS Tagging dan tanpa menggunakan PoS Tagging [17].

**2.5.1. Menggunakan Pos Tagging**

Untuk menghitung melakukan perhitungan dengan menggunakan Delta TF-IDF dapat dilakukan dengan menggunakan formula 1.

$$tf - id_{anak,D1} = 2 * \log_2 \left( \frac{1+0,5}{1+1+0,5} \right) = 2 * 0 = 0 \quad (1)$$

Pada perhitungan diatas  $N_p = 1$  dimana total dokumen kelas positif adalah 1 dan hal tersebut juga berlaku untuk  $N_n$ . Selanjutnya Nilai A didapatkan dari total kemunculan kata anak dalam D1 (yang berlaku sebagai kelas positif) dengan kemunculan lebih dari 0 (dianggap 1) dan Nilai C didapatkan total kemunculan kata anak dalam D3 (yang berlaku sebagai kelas negative) dengan kemunculan lebih dari 0 (dianggap 1) [18].

**2.5.2. Tanpa Menggunakan Pos Tagging**

Untuk menghitung melakukan perhitungan dengan menggunakan Delta TF-IDF dapat dilakukan dengan menggunakan formula 2.

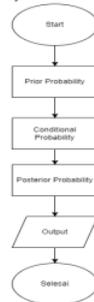
$$tf - id_{anak,D1} = 2 * \log_2 \left( \frac{1+0,5}{1+1+0,5} \right) = 2 * 0 = 0 \quad (2)$$

Pada perhitungan diatas  $N_p = 1$  dimana total dokumen kelas positif adalah 1 dan hal tersebut juga berlaku untuk  $N_n$ . Selanjutnya Nilai A didapatkan dari total kemunculan kata anak dalam D1 (yang berlaku sebagai kelas positif) dengan kemunculan lebih dari 0 (dianggap 1) dan Nilai C didapatkan total kemunculan kata anak dalam D3 (yang berlaku sebagai kelas negative) dengan kemunculan lebih dari 0 (dianggap 1).

**2.5.3. Klasifikasi Dengan Naïve Bayes dan Random Forest**

**a) Klasifikasi Naive Bayes**

Berikut pada Gambar 2 merupakan tahapan yang digunakan klasifikasi naïve bayes.



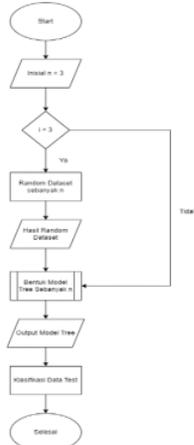
**Gambar 2.** Tahapan Naive Bayes

Berdasarkan Gambar 2 merupakan Tahapan Naïve Bayes, sebagai berikut [19]:

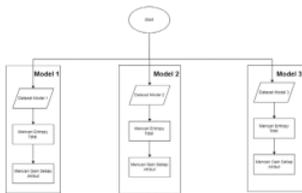
- 1) Prior Probability (Probabilitas Prior): Tahap pertama adalah menghitung probabilitas prior, yang merupakan probabilitas munculnya suatu kategori atau kelas sebelum adanya informasi dari fitur-fitur yang ada. Ini memberikan gambaran awal tentang seberapa sering suatu kategori muncul di dalam data.
- 2) Conditional Probability (Probabilitas Bersyarat): Setelah probabilitas prior dihitung, langkah berikutnya adalah menghitung probabilitas bersyarat untuk setiap fitur dalam setiap kategori. Probabilitas ini

menunjukkan seberapa mungkin suatu fitur akan muncul jika diketahui bahwa data termasuk dalam suatu kategori tertentu.

- 3) Posterior Probability (Probabilitas Posterior): Tahapan berikutnya adalah menghitung probabilitas posterior, yang menggabungkan informasi dari probabilitas prior dan probabilitas bersyarat. Probabilitas posterior memberikan perkiraan seberapa mungkin suatu data masuk ke dalam suatu kategori berdasarkan fitur-fitur yang diamati.
  - 4) Output: Langkah terakhir pada flowchart Naive Bayes adalah menghasilkan output atau prediksi berdasarkan probabilitas posterior. Data diklasifikasikan ke dalam kategori dengan probabilitas posterior tertinggi, yang merupakan prediksi atau output akhir dari algoritma Naive Bayes.
- b) Klasifikasi Random Forest
- Random Forest merupakan salah satu metode machine learning yang dimanfaatkan untuk klasifikasi data yang memiliki jumlah banyak [20]. Berikut pada Gambar 3 merupakan flowchart dari random forest dan Gambar 3 merupakan flowchart tree.



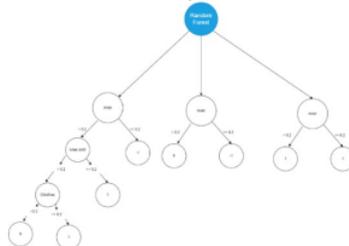
Gambar 3. Flowchart Random Forest



Gambar 4. Flowchart Tree

Berdasarkan Gambar 3 dan Gambar 4 sebagai berikut:

- 1) Inisial N  
Dari Gambar 3 hingga Gambar 4, dapat dijelaskan bahwa *random forest* merupakan sekumpulan *tree*. Jika *random forest* akan membentuk 3 model *tree* dikarenakan  $N = 3$ . ( $K$ ) dimana untuk membentuk model *tree* akan menggunakan formula *decision tree*, maka *random forest* akan melakukan pengacakan (*random*) dataset untuk 3 model dimana setiap model memiliki dataset nya tersendiri. Sehingga *flowchart* yang terbentuk dapat dilihat pada Gambar 5.
- 2) *Random Dataset* Sebanyak  $N$   
Pengambilan akan dilakukan secara *random* dengan kemungkinan terdapat kemunculan data yang pada hasil tiap bootstrap dan bahkan tidak sama sekali. Pengambilan disini akan menggunakan index data untuk memudahkan pembacaan alur.
- 3) Mencari *Entropy* Total  
Untuk menghitung *entropy* perlu mengetahui probabilitas label terhadap total kasus. Pada dataset ini memiliki 2 label yaitu 1 dan -1.
- 4) Mencari *Gain* Setiap Atribut  
Berikut merupakan contoh perhitungan model ke-1. Untuk mencari *gain*, diperlukan mencari *entropy* setiap fitur terhadap setiap kriteria pada fitur tersebut dengan menggunakan persamaan (2.6). Contoh pada fitur pada anak terdapat 2 kriteria  $< 0.2$  dan  $\geq 0.2$  maka perlu diketahui total kemunculan setiap kriteria terhadap terhadap label.
- 5) Output  
Dari Information Gain di atas didapatkan model *tree* yang nantinya akan dilakukan penggabungan menjadi sebuah forest. Jika divisualkan menjadi Gambar 5.



Gambar 5. Visual Random Forest

#### 2.5.4. Prediksi Data Test

Pada hasil *Random Forest* diatas, didapatkan bahwasanya setiap model *tree* memiliki 1 akar dan 2 leaf (daun) pada model *tree* dan 2 dan 3, namun berbeda dengan model *tree* 1. Hal ini dikarenakan distribusi data yang tidak merata sehingga harus melakukan iterasi hingga menemukan distribusi yang normal. Dengan setiap model *tree* memiliki *feature root* dan *leaf* yang berbeda. Untuk menentukan kelas yang akan diprediksi akan menggunakan sistem **majorit vactory** yang dimana jika 2 model memprediksi 1 dan 1 model memprediksi 0 maka hasil prediksi data testing adalah 1.

2.5.5. Evaluasi

Setelah dilakukan klasifikasi menggunakan algoritma Random Forest dan Naïve Bayes selanjutnya dilakukan proses evaluasi dengan menggunakan metode confusion matrix. Confusion Matrix merupakan sebuah tabel yang memberikan informasi perbandingan hasil klasifikasi yang dilakukan oleh sistem (prediksi) dengan hasil klasifikasi yang sebenarnya [21]. Pengukuran kinerja menggunakan confusion matrix, terdapat 4 istilah diantaranya yaitu sebagai berikut [22]:

- a) *True Positive* (TP) yaitu jumlah data positif yang terklasifikasi dengan benar oleh sistem.
- b) *True Negative* (TN) yaitu jumlah data negatif yang terklasifikasi dengan benar oleh sistem.
- c) *False Positive* (FP) yaitu jumlah data positif namun terklasifikasi salah oleh sistem.
- d) *False Negative* (FN) yaitu jumlah data negatif namun terklasifikasi salah oleh sistem.

3. HASIL

Hasil penelitian ini menjelaskan perbandingan kinerja metode random forest dan naive bayes untuk analisis sentimen masyarakat terhadap childfree pada twitter.

A. Hasil Implementasi Metode Naïve Bayes Dan Random Forest

Dalam implementasi metode *naïve bayes* dan *random forest* dilakukan menggunakan bahasa pemrograman *Python* untuk proses klasifikasi pada dataset *childfree* pada *Twitter* [23] Dataset penelitian ini berupa pengambilan data dari *Twitter* yang berisi tweet opini masyarakat terkait *childfree*. Data yang digunakan dalam penelitian ini berjumlah 1309 *tweet* yang diambil dari tanggal 26 Februari 2023 hingga 8 Maret 2023. Dataset yang telah dikumpulkan dilakukan *load* dataset untuk melihat data tersebut yang hasilnya dapat dilihat pada Tabel 1.

Tabel 1. Read Dataset

Id	name	web	Created_at	username	tweet	location
0	1633380776996904961	lhaiki.lho	2023-03-08 08:14:58+00:00	lhaiki.lho	Childfree? anak alian gak bakal ngerasa in nyuruh ...	NaN
1	1633372857848991749	Kore of Cht honia	2023-03-08 07:43:30+00:00	satur nom eus	Happy Women's Day untuk single mother s, peremp ...	NaN
2	1633367119059623936	Ha bibi Ha bib ah	https://t.c o/Yr7Hq vBYJ 07:20:42+00:00	ibnu rozii	@DitjenPajak RI Hufti mungkin skrng dipikir any...	Depan Lanyar

3	1633353822914609152	Dor r	2023-03-08 06:27:52+00:00	NaN	gak enna	@detikcom Sepandai apapun kalian menyipban...	NaN
4	1633343728181198848	-	2023-03-08 05:47:45+00:00	NaN	vnylp	@Askrifess Zalim bgt ibunya, yang begini begini...	NaN
04	162971350458553360384	Ajeng Kar tika N	2023-02-26 07:49:23+00:00	NaN	ajengggkk artika	Aku emak anak 3 mlah ngwa st nti Javawanti teme...	Mala ng, East Java
1305	1629748585041625094	gitrlfrt nd Expe riences	https://t.c o/jp2sdd EJQp 02-26 07:41:56+00:00	2023-02-26 07:41:56+00:00	Cha caV arma	Kalian terlalu sibuk sama tantrum dan malut ja...	Up per ear th
1306	1629748334561996800	rosaa	https://t.c o/HFZP9vdH 02-26 07:40:56+00:00	2023-02-26 07:40:56+00:00	psiholic	@zhafiragyla Sender nya kepo bgt bjir, org baru...	Unive rsitas Gadjah Mada
1307	1629745933473955840	Sapiens	2023-02-26 07:31:24+00:00	NaN	Hom o_sapien s21	Twit ini kejam karena survival of the fittest ...	NaN
1308	1629745850334208000	Yokan C. A.	https://t.c o/e9dYqt IVO 02-26 07:31:04+00:00	2023-02-26 07:31:04+00:00	Yok an g...	@Greesecl JK T48 kacing mu gak childfree cel?	NaN

Dataset tersebut dilakukan *preprocessing* yang terdiri dari *case folding*, *tokenizing*, *filtering* dan *stemming* tanpa menggunakan *pos tagger* yang hasilnya dapat dilihat pada Tabel 2.

Tabel 2. Hasil Preprocessing

user name	tweet	case folding	tokenizing	filtering	stemming	label
0	lhaiki_lho	Childfree? anak bakal ngerasa in	childfree? anak bakal ngerasa in	[child free, kalia ngera n, nyur uh, anak	child free ngerasai n	Neutral

Commented [U1]: Sudah diperbaiki untuk semua tabel sesuai revisian kemudian model dari semua tabel terbuka tanpa gar

	nyuruh ...	nyuruh a...	ngera sain, n...	matii n kran in kran ai...	anak mati in kran ai...	
1	satum omeus	Happy Women' s Day untuk single mothers - peremp. ...	happy women s day, untuk single mothers - peremp. ...	[happ y, wom ens, day, untuk singl e moth ers pere mpua n foku. ...	happ y wom ens day singl e mot hers per mpu an foku ...	Pos itiv e
2	iburo zii	@Dijen PajakRI Huftt mungkin n skmg dipikira ny...	huftt mungk in skmg dipikir annya childfr ee ti...	mung kin, sekar ang, dipiki ranya a, child. ...	dipik iran ya child free buru k	Ne gati ve
3	ggake nna	@detik com Sepanda i apapun kalian menyim pan ban...	sepan dai apapun kalian menyi mpan bangka i cepa...	[sepa ndai, apap un meny impa n, bang kai cepat lamb at...]	pan dai apa simp an bang kai cepa t lamb at aro ma b...	Ne gati ve
4	vnylp	@Askrl fess Zalim bgt ibunya, yang begini ini nih begin...	zalim bgt ibunya yang begini ini nih y...	[zali m, begit u, ibuny a, yang begin i, begin i, ...]	zali m ibu child free	Ne utra l
1 2 4 8	ajengg kkarti ka	Aku emak emak anak 3 malah ngewant i wanti teme...	aku emak emak malah ngewa nti wanti teme...	[aku, emak emak emak anti wanti tema n dek et ku bi...]	ema k ema k anak nge want i want i tema n deke tku bi...	Ne utra l
1 2 4 9	Chaca Vam a	Kalian terlalu sibuk sama tantrum dan	kalian terlalu sibuk sama tantru m dan	[kalia n, tantr um mulu t jahat - sama, gitsa	sibu k tantr um mul ut jahat	Ne gati ve

mulut ja...	mulut ja...	tantru m, dan, m...	v lupa popul eri...	gitsa v lupa popu leri...
----------------	----------------	------------------------------	------------------------------	---------------------------------------

Dari Tabel 3 hasil *preprocessing* data, dilakukan pengelompokan jumlah label dari setiap label. Dalam penelitian ini menggunakan 3 jenis label yaitu Negative dengan total data 331, Neutral berjumlah 627 data dan Positive berjumlah 295 data. Selanjutnya, dilakukan pembagian dataset menjadi 2 jenis data yaitu data *train* 80 % dan data *test* 20 % yang dapat dilihat pada Tabel 3 dan Tabel 4. Pada data *train* berjumlah 1002 data dan data *test* berjumlah 251 data.

Tabel 3. Data Training

stem	label	
0	liat bawa childfree	Neutral
1	kampanye childfree dosa rendah anak indah utam...	Negative
2	cocok childfreetanpa kondomanti tanggal merah	Positive
3	ohh alas marak childfree	Neutral
4	childfree prolife tunggu lahir buang	Positive
...	...	...
997	endeavor bunda rei childfree husbu gwe	Neutral
998	kucing anut childfree	Positive
999	teh rafael mending mah childfree	Neutral
1000	awet muda childfree	Positive
1001	gilir kayak gin gamau dukung dilakuin ortu bab...	Negative

Tabel 4. Data Testing

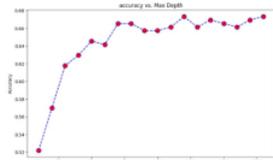
stem	label	
0	childfree gitasav tunggu pensi	Positive
1	duhh bingung pikir childfree gitasav tolol kay...	Negative
2	minimal pikirin childfree	Neutral
3	allah jahat anak someday against orang milih c...	Negative
4	andai gitasav hidup paradisi kampanye childfree	Neutral
...	...	...
246	salah gitasav ngomong childfree orang juang be...	Negative
247	ngebahas childfree kadang nyelutuk hah ngapain...	Negative
248	suruh childfree anak tega	Neutral
249	betul berat orgtua pantesan childfree liat sis...	Negative
250	tuju childfree saiki goblok arek enom	Neutral

Pada masing-masing data dilakukan pembobotan kata menggunakan TF-IDF tradisional dengan memasukkan nilai delta. Dalam konteks ini, delta mengacu pada perubahan atau perbedaan dalam frekuensi relatif suatu kata antara dua dokumen atau lebih, yang hasilnya dapat dilihat pada Tabel 5 dan Tabel 6.

Tabel 5. Hasil Delta TFIDF Train Data

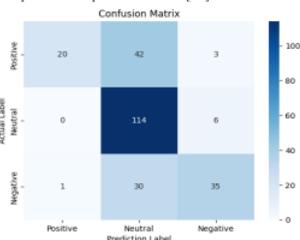






Gambar 7. Hasil Visualisasi Plot Metode Random Forest

Kemudian dilakukan evaluasi menggunakan *confusion matrix* yang didapatkan hasil pada Gambar 16[24].



Gambar 8. Confusion Matrix Metode Random Forest

Pada Gambar 8 dapat dijelaskan bahwa metode *random forest* tanpa *pos tagger* diperoleh hasil prediksi label *Positive* berjumlah 65 data yang diuji, terdapat 20 data yang terklasifikasi dengan benar dan terdapat kesalahan prediksi berjumlah 42 data yang masuk kedalam label *Neutral* dan 3 data masuk kedalam label *Negative*. Sedangkan, pada label *Neutral* dengan jumlah data 120 yang diuji, terdapat 114 data yang terklasifikasi dengan benar, kesalahan prediksi berjumlah 16 data pada label *Negative*. Pada label *Negative* dengan jumlah data 66 data yang diuji, terdapat 35 data yang terklasifikasi dengan benar, terdapat kesalahan prediksi berjumlah 1 data yang masuk kedalam label *Positive* dan 30 data masuk kedalam label *Neutral*.

Pada hasil evaluasi metode *random forest* tanpa menggunakan *pos tagger* mendapatkan nilai *accuracy* 67,33 %, *precision* sebesar 74,88 %, *recall* sebesar 67,33 % dan *F-Score* sebesar 64,4 %. Hasil perhitungan *confusion matrix* pada metode *random forest* tanpa *pos tagger* dapat dilihat pada Tabel 9 dan Tabel 10[25].

Tabel 9. Hasil Confusion Matrix Metode Random Forest Tanpa Pos Tagger

Accuracy	67.33 %
Precision	74.88 %
Recall	67.33 %
F-Score	64.4 %

Tabel 10. Hasil Klasifikasi Metode Random Forest

stem	label	pred_label_n	pred_label_r
0	childfree gitasav	Positive	Positive
		Positive	Neutral

<https://doi.org/10.25077/TEKNOSI.xxxxx>

	tunggu pensi duhh bingung pikir childfree gitasav tolol kay...	Negative	Neutral	Negative
2	minimal pikirin childfree	Neutral	Neutral	Neutral
3	allah jahat anak someday against orang milih c...	Negative	Neutral	Negative
4	andai gitasav hidup paradisi kampanye childfree	Neutral	Neutral	Neutral
24	salah gitasav ngomong childfree orang juang be...	Negative	Negative	Negative
24	ngebahas childfree kadang nyelutuk hah ngapain...	Negative	Negative	Positive
24	suruh childfree anak tega	Neutral	Neutral	Neutral
24	betul berat ortua pantasan childfree liat sis...	Negative	Negative	Negative
25	tujuan childfree saiki goblok arek enom	Neutral	Neutral	Neutral

Tabel 11. Hasil Perbandingan Evaluasi Metode Naïve Bayes Random Forest

		Naive Bayes (%)	Random Forest (%)
0	Accuracy	69.72	67.33
1	Precision	71.86	74.88
2	Recall	69.72	67.33
3	F1-score	68.94	64.40

Pada Tabel 11 merupakan hasil perbandingan evaluasi metode *naive bayes* dan *random forest* tanpa *pos tagger* yang mendapatkan hasil bahwa *accuracy*, *recall* dan *f1-score* metode

*naïve bayes* lebih unggul dari metode *random forest*. Sedangkan, pada nilai *precision* metode *random forest* lebih unggul dari metode *naïve bayes*.

**B. Hasil Implementasi Program**

Pada tahap ini dilakukan implementasi algoritma Naive Bayes dan Random Forest ke dalam bentuk website untuk menampilkan Hasil analisis sentimen masyarakat terhadap opini di Twitter.

**1. Implementasi Hasil Analisis Sentimen**

Implementasi aplikasi tampilan analisis data sentimen, lakukan import data dari hasil scrapping untuk mengetahui hasil analisis sentimeen dengan metode *naive bayes* dan *random forest*. Berikut Tampilan Import file, dapat dilihat pada Gambar 9.



**Gambar 9.** Tampilan Import File

**2. Implementasi Hasil Halaman Detail Data**

Halaman ini akan menampilkan seluruh atribut dan informasi yang terkait dengan entri tersebut, seperti teks tweet, label sentimen, dan atribut lainnya, dapat dilihat pada Gambar 10.



**Gambar 10.** Tampilan Detail Data

**3. Implementasi Update Tabel**

Dalam melakukan update label pada sistem pengguna dapat melakukan secara manual, dimana pengguna dapat memilih label positif, negatif, atau netral. Berikut tampilan update label dapat dilihat pada Gambar 11.



**Gambar 11.** Tampilan Update Tabel

**4. Implementasi Hasil Klasifikasi Menggunakan Data Training dan Testing**

Halaman Hasil Klasifikasi Menggunakan Data Training dan Testing merupakan bagian penting dalam proses analisis data dan evaluasi model. Halaman ini menyajikan hasil klasifikasi yang diperoleh dari model

yang telah dilatih menggunakan data training dan diuji dengan data testing, dapat dilihat pada Gambar 12.



**Gambar 12.** Tampilan Hasil Klasifikasi

**5. Implementasi Hasil Klasifikasi Menggunakan Data Baru**

Halaman Hasil Klasifikasi Menggunakan Dataset Baru adalah tampilan yang bertujuan untuk menampilkan hasil dari klasifikasi data yang telah dianalisis menggunakan model yang telah dilatih sebelumnya, namun dengan dataset yang baru, dapat dilihat pada Gambar 13.



**Gambar 13.** Tampilan Hasil Dataset Baru

**C. Pengujian Hasil Pelabelan**

Pengujian hasil pelabelan melibatkan analisis sentimen dengan dua teknik pelabelan dataset yang berbeda, yakni menggunakan corpus dan pelabelan manual oleh manusia. Dalam pelabelan menggunakan corpus dengan algoritma *naïve bayes*, terdapat akurasi sebesar 71.7%, Precision = 76.32%, Recall = 71.7%, dan F-Score = 70.89%. Sementara itu, algoritma *random forest* menghasilkan akurasi sebesar 81.13%, Precision = 82.73%, Recall = 81.137%, dan F-Score = 80.54%. Dalam pelabelan manual menggunakan algoritma *naïve bayes*, terdapat akurasi sebesar 69.81 %, Precision = 75.3 %, Recall = 69.81 %, F-Score = 67.92 %. Untuk hasil akurasi menggunakan *random forest*, terdapat nilai akurasi sebesar 79.25 %, Precision = 82.12 %, Recall = 79.25 %, F-Score = 78.99 %%. Berikut hasil akurasi menggunakan corpus dan manual, dapat dilihat pada Tabel 12 hingga 13.

**Tabel 12.** Hasil *Confusion Matrix* Metode *Naive Bayes* Dan *Random Forest*

	Accuracy	Precision	Recall	f-score
<b>Naïve Bayes</b>				
Tanpa Pos Tagger	69.72%	71.86%	69.72%	68.94%
Menggunakan Pos Tagger	78.76%	79.00%	78.76%	78.02%
<b>Random Forest</b>				
Tanpa Pos Tagger	67.33%	74.88%	67.33%	64.40%
Menggunakan Pos Tagger	81.85%	83.66%	81.85%	81.65%

**Tabel 13.** Perbandingan Pelabelan Manual dan Menggunakan *Corpus*

Pelabelan <i>Corpus</i>		Pelabelan Manual	
<i>Random Forest</i>	<i>Naïve Bayes</i>	<i>Random Forest</i>	<i>Naïve Bayes</i>
79.25%	69.81%	81.13%	71.70%

#### 4. PEMBAHASAN

Dalam penelitian ini, analisis sentimen masyarakat terhadap opini di *Twitter* telah dilakukan. Data yang digunakan terdiri dari 1309 *tweet* yang dikumpulkan dari tanggal 26 Februari 2023 hingga 8 Maret 2023. Data tersebut dibagi menjadi dua bagian: data *training*, yang digunakan untuk melatih model analisis sentimen, dan data *testing*, yang digunakan untuk menguji performa model. Terdapat enam sampel data yang mewakili berbagai kelas, dan dataset ini dibagi menjadi dua bagian, yaitu *train* (80%) dan *test* (20%). Hasil evaluasi menggunakan *Pos Tagger* antara metode *Naïve Bayes* dan *Random Forest* terhadap opini masyarakat di *Twitter* menunjukkan bahwa *Random Forest* memiliki akurasi yang lebih tinggi, yaitu 81.85%, dibandingkan dengan *Naïve Bayes* yang memiliki akurasi sebesar 78.75%. Artinya, *Random Forest* berhasil mencapai akurasi yang lebih tinggi dalam mengklasifikasikan sentimen masyarakat terhadap opini di *Twitter* dibandingkan dengan metode *Naïve Bayes*. Sementara hasil evaluasi tanpa *Pos Tagger* menggunakan *naïve bayes* dan *random forest*, menghasilkan nilai akurasi sebesar 69.72% untuk *Naïve Bayes*, dan nilai akurasi *Random Forest* sebesar 67.33%. Artinya, *Naïve Bayes* berhasil mencapai akurasi yang lebih tinggi dalam mengklasifikasikan sentimen masyarakat terhadap opini di *Twitter* dibandingkan dengan metode *Random Forest*.

#### 5. KESIMPULAN

Berdasarkan hasil penelitian mengenai sentimen masyarakat terhadap 'childfree' di media sosial *Twitter*, dapat disimpulkan bahwa pengumpulan data dilakukan secara efektif dengan menggunakan teknik *scrapping* dan format *.csv* berbasis lokasi. Penelitian ini berhasil mengimplementasikan metode *Random Forest* dan *Naïve Bayes* dengan menggunakan bahasa pemrograman *python* dengan memanfaatkan beberapa *library python* seperti *numpy*, *pandas*, *sklearn.naive\_bayes*, *sklearn.ensemble.RandomForestClassifier*, *sklearn.metrics* serta beberapa *library* pendukung lainnya sehingga dapat membantu dalam *Random Forest* dan *Naïve Bayes*, untuk melakukan analisis sentimen terhadap mengimplementasikan algoritma *Random Forest* dan *Naïve Bayes* dalam melakukan analisis sentimen terhadap topik "childfree" di media sosial *Twitter*. Pada penelitian ini evaluasi dilakukan dengan menggunakan metode *Confusion Matrix* atau sebuah tabel yang memberikan informasi perbandingan hasil klasifikasi yang dilakukan oleh sistem (prediksi) dengan hasil klasifikasi yang sebenarnya. Hasil evaluasi menunjukkan bahwa dalam kasus ini, *Random Forest* memiliki akurasi yang lebih tinggi (81.85%) dibandingkan dengan *Naïve Bayes* (78.75%) ketika menggunakan *Pos Tagger*. Namun, tanpa menggunakan *Pos Tagger*, *Naïve Bayes* mencapai akurasi lebih tinggi (69.72%) daripada *Random Forest* (67.33%). Ini mengindikasikan bahwa pemilihan algoritma dapat dipengaruhi oleh aspek seperti penggunaan *Pos Tagger* dalam pengolahan data. Sehingga, keputusan pemilihan algoritma

<https://doi.org/10.25077/TEKNOSI.xxxxx>

sebaiknya dipertimbangkan dengan cermat tergantung pada metode pengolahan data yang digunakan dalam analisis sentimen *childfree* di media sosial *Twitter*.

#### DAFTAR PUSTAKA

- [1] V. Cornelia, N. Sugianto, N. Glori, and M. Theresia, "Fenomena Childfree dalam Perspektif Utilitarianisme dan Eksistensialisme," *JPraxis J. Filsafat Terap.*, vol. 1, no. 1, pp. 1–16, 2022, doi: 10.11111/moderasi.xxxxxx.
- [2] S. Al Usman, M. Maksudin, N. Nurhadi, T. Thonthowi, and M. A. D. Insan Labib, "Childfree and the Position of Children From an Islamic Perspective in Indonesia," *Huk. Islam*, vol. 23, no. 1, p. 16, 2023, doi: 10.24014/jhi.v1i23.20360.
- [3] R. P. N. Sari, Y. N. Nobisa, J. M. Sali, I. Iskandar, B. K. Paradila, and A. S. Rahman, "Pandangan Tokoh Muhammadiyah Di Kota Kupang Terhadap Childfree," *UluMuddin J. Ilmu-ilmu Keislaman*, vol. 12, no. 2, pp. 357–372, 2022, doi: 10.47200/ulumuddin.v12i2.1336.
- [4] M. I. Fikri, T. S. Sabrila, Y. Azhar, and U. M. Malang, "Comparison of the Naïve Bayes Method and Support Vector Machine on Twitter Sentiment Analysis," *SMATIKA J. STIKI Inform. J.*, vol. 10, no. 2, pp. 71–76, 2020.
- [5] A. S. Rahayu, A. Fauzi, and R. Rahmat, "Komparasi Algoritma Naïve Bayes Dan Support Vector Machine (SVM) Pada Analisis Sentimen Spotify," *J. Sist. Komput. dan Inform.*, vol. 4, no. 2, p. 349, 2022, doi: 10.30865/json.v4i2.5398.
- [6] P. Karthika, R. Mungeswari, and R. Manoranjithem, "Sentiment Analysis of Social Media Network Using Random Forest Algorithm," *IEEE Int. Conf. Intell. Tech. Control. Optim. Signal Process. INCOS 2019*, pp. 1–5, 2019, doi: 10.1109/INCOS45849.2019.8951367.
- [7] S. Surohman, S. Aji, R. Rousyati, and F. F. Wati, "Analisa Sentimen Terhadap Review Fintech Dengan Metode Naïve Bayes Classifier Dan K-Nearest Neighbor," *EVOLUSI J. Sains dan Manaj.*, vol. 8, no. 1, pp. 93–105, 2020, doi: 10.31294/evolusi.v8i1.7535.
- [8] A. S. Putra, D. Anubhakti, and L. L. Hin, "APPLICATION OF K-NEAREST NEIGHBOR ALGORITHM FOR SENTIMENT ANALYSIS OF TWITTER USERS ON ELECTRIC VEHICLE," pp. 736–744.
- [9] E. Fitri, "Analisis Sentimen Terhadap Aplikasi Ruangguru Menggunakan Algoritma Naïve Bayes, Random Forest Dan Support Vector Machine," *J. Transform.*, vol. 18, no. 1, p. 71, 2020, doi: 10.26623/transformatika.v18i1.2317.
- [10] N. Mingkase and I. Rohmaniyah, "Konstruksi gender dalam problematika childfree di sosial media *Twitter*," *Yinyang J. Stud. Islam Gen. dan Anak*, vol. 17, no. 2, pp. 201–222, 2022, doi: 10.24090/yinyang.v17i2.6486.
- [11] T. Fadiyah Basar, D. E. Ratnawati, and I. Arwani, "Analisis Sentimen Pengguna *Twitter* terhadap Pembayaran Cashless menggunakan *Shopeepay* dengan Algoritma *Random Forest*," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 6, no. 3, pp. 1426–1433, 2022.
- [12] F. Rahutomo, P. Y. Saputra, and M. A. Fidyawan, "Implementasi *Twitter* Sentiment Analysis Untuk Review Film Menggunakan Algoritma Support Vector Machine," *J. Inform. Polinema*, vol. 4, no. 2, p. 93, 2018, doi:

10.33795/jip.v4i2.152.

[13] Y. Pratama, D. T. Mardiansyah, and K. M. Lhaksmana, "Analisis Sentimen Kendaraan Listrik Pada Media Sosial Twitter Menggunakan Algoritma Logistic Regression dan Principal Component Analysis," *J. Media Inform. Budidarma*, vol. 7, no. 1, pp. 529-535, 2023, doi: 10.30865/mib.v7i1.5575.

[14] M. Z. Nafan and A. E. Amalia, "Kecenderungan Tanggapan Masyarakat terhadap Ekonomi Indonesia berbasis Lexicon Based Sentiment Analysis," *J. Media Inform. Budidarma*, vol. 3, no. 4, p. 268, 2019, doi: 10.30865/mib.v3i4.1283.

[15] M. Ojeda-Hernández, D. López-Rodríguez, and Á. Mora, "Lexicon-based sentiment analysis in texts using Formal Concept Analysis," *Int. J. Approx. Reason.*, vol. 155, pp. 104-112, 2023, doi: 10.1016/j.ijar.2023.02.001.

[16] A. M. Pravina, I. Cholissodin, and P. P. Adikara, "Analisis Sentimen Tentang Opini Maskapai Penerbangan pada Dokumen Twitter Menggunakan Algoritme Support Vector Machine (SVM)," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 3, no. 3, pp. 2789-2797, 2019.

[17] C. Amalia and Y. Sibaroni, "Analisis Sentimen Data Tweet Menggunakan Model Jaringan Saraf Tiruan Dengan Pembobotan Delta TF-idf," *eProceedings ...*, vol. 7, no. 2, pp. 7810-7820, 2020.

[18] I. Firmansyah *et al.*, "Klasifikasi Kelas Kata ( Part-of-Speech Tagging ) Untuk Bahasa Classification of Words Class ( Part-of-Speech Tagging ) for Bahasa Madura Using Viterbi," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 8, no. 5, 2021, doi: 10.25126/jtiik.202184483.

[19] Buttercup, "Naive Bayes Classifier," 2021.

[20] A. R. Basar and E. Safira, "Rancang Bangun Dashboard Untuk Monitoring Inventory Barang Pada Perusahaan T-Obenk Kaos Batam," *JR J. Responsive Tek. Inform.*, vol. 4, no. 2, pp. 51-59, 2020, doi: 10.36352/jr.v4i2.179.

[21] U. D. Nuswantoro, "Pertemuan Ke 8."

[22] D. Susanti and T. Wahyuni, "Analisis Potensi Bencana Alam Tanah Longsor Kabupaten Majalengka Menggunakan Algoritma Naive Bayes Classifier," *INFOTECH J.*, vol. 9, no. 2, pp. 299-306, 2023, doi: 10.31949/infotech.v9i2.5645.

[23] J. Florensius Sianipar, Y. R. Ramadhan, and I. Jaelani, "Analisis Sentimen Pembangunan Kereta Cepat Jakarta-Bandung di Media Sosial Twitter Menggunakan Metode Naive Bayes," *Media Online*, vol. 4, no. 1, pp. 360-367, 2023, doi: 10.30865/klk.v4i1.1033.

[24] P. A. W. RD, I. Susilawati, and A. Witanti, "Analisis Sentimen pada Komentar Aplikasi MyPertamina dengan Metode Multinomial Naive Bayes," *Informatics Artif. ...*, vol. 1, no. 1, 2023.

[25] N. Puspitasari, A. Septiarini, and A. R. Aliudin, "Metode K-Nearest Neighbor Dan Fitur Warna Untuk Klasifikasi Daun Sirih Berdasarkan Citra Digital," *PROSISKO J. Pengemb. Ris. dan Obs. Sist. Komput.*, vol. 10, no. 2, pp. 165-172, 2023, doi: 10.30656/prosisko.v10i2.6924.

## NOMENKLATUR

$N_p$	total dokumen kelas positif
$N_n$	total dokumen kelas negatif
Nilai A	total kemunculan kata anak dalam D1
Nilai C	total kemunculan kata anak dalam D3
root	simpul yang terletak paling atas atau biasa disebut sebagai akar dari pohon keputusan
leaf	simpul terakhir yang hanya memiliki satu input dan tidak memiliki output
x	berisi matriks atau array yang mengandung atribut yang digunakan untuk melakukan prediksi terhadap variabel dependen (y)
y	berupa array atau vektor yang memuat nilai yang ingin diprediksi atau dijelaskan oleh fitur yang disimpan dalam variabel x
X_train	Variabel ini digunakan untuk menyimpan dataset fitur yang akan digunakan untuk melatih model dalam pembelajaran mesin
X_test	Variabel ini berfungsi untuk menyimpan dataset fitur yang akan digunakan untuk menguji model yang telah dilatih
y_train	berisi nilai yang sesuai dengan setiap baris data dalam X_train, yang akan digunakan sebagai dasar untuk melatih model
y_test	berfungsi untuk mengevaluasi seberapa baik model yang dilatih dapat melakukan prediksi pada data baru yang belum pernah dilihat selama pelatihan
train_data	biasanya digunakan untuk menyimpan data yang akan digunakan untuk melatih model dalam pembelajaran mesin atau analisis data
train_negatif	Variabel ini digunakan untuk menyimpan data atau sampel teks yang diklasifikasikan sebagai negatif dalam analisis sentimen
train_netral	Variabel ini digunakan untuk menyimpan data atau sampel teks yang diklasifikasikan sebagai netral dalam analisis sentimen
train_positif	Variabel digunakan untuk menyimpan data atau sampel teks yang diklasifikasikan sebagai positif dalam analisis sentimen

Commented [U2]: Sudah ditambahkan untuk nomenklaturnya sesuai dengan contoh

Nomenklatur: semua variable dituliskan di nomenkla

## BIODATA PENULIS



(Sevima)

Giwa Salsabila Ferari Irianto  
Sedang belajar sebagai mahasiswa prodi Teknik Informatika, Fakultas Teknik, Universitas 17 Agustus 1945 Surabaya. Saat ini juga bekerja sebagai Customer Onboarding Specialist, Surabaya, East Java di PT Sentra Vidya Utama



Fajar Astuti Hermawati  
Bergabung di Prodi Teknik Informatika, Fakultas Teknik, Universitas 17 Agustus 1945 Surabaya sebagai staf pengajar atau dosen sejak tahun 2000. Lulus doctoral dari prodi S3 Ilmu Komputer Institut Teknologi Sepuluh Nopember Surabaya tahun 2019. Bidang riset yang diteliti meliputi kecerdasan artifisial, pengolahan citra digital, pengenalan pola, computer vision dan data mining serta NLP.

<https://doi.org/10.25077/TEKNOSI.xxxxx>

**LAMPIRAN**

-

# Giwa

---

## ORIGINALITY REPORT

---

12%

SIMILARITY INDEX

12%

INTERNET SOURCES

8%

PUBLICATIONS

6%

STUDENT PAPERS

---

## PRIMARY SOURCES

---

1	Submitted to Universitas Andalas Student Paper	4%
2	pdfs.semanticscholar.org Internet Source	1%
3	journal.forikami.com Internet Source	1%
4	dspace.uii.ac.id Internet Source	1%
5	media.neliti.com Internet Source	1%
6	doaj.org Internet Source	1%
7	ojs.unimal.ac.id Internet Source	1%
8	journals.usm.ac.id Internet Source	1%
9	ojs.unud.ac.id Internet Source	1%

---

10

j-ptiik.ub.ac.id  
Internet Source

1 %

---

Exclude quotes      On

Exclude matches      < 1%

Exclude bibliography      On