Analisis Sentimen Opini Publik Terhadap Fenomena *Childfree* Menggunakan Algoritma *Naïve Bayes* Pada Media Sosial *Twitter*

1st Salma Haniyah
Fakultas Rekayasa Industri
Universitas Telkom
Bandung, Indonesia
mailto:salmahaniyah@student.telkomuniversity.ac.id

2nd Deden Witarsyah
Fakultas Rekayasa Industri
Universitas Telkom
Bandung, Indonesia
mailto:dedenw@telkomuniversity.ac.id

3rd Edi Sutoyo
Fakultas Rekayasa Industri
Universitas Telkom
Bandung, Indonesia
mailto:edisutoyo@telkomuniversity.ac.id

Abstrak **- perkembangan media sosial, terutama *twitter*, sebagai sarana komunikasi dan penyaluran pendapat semakin pesat. Gerakan *childfree*, yakni keputusan individu untuk tidak memiliki anak, telah menjadi topik yang ramai dibicarakan di media sosial. Hal ini mencerminkan perubahan pandangan masyarakat terhadap kehidupan keluarga dan memiliki dampak signifikan pada dinamika sosial dan kebijakan publik. Penelitian ini bertujuan untuk menganalisis sentimen pengguna *twitter* terhadap gerakan *childfree* menggunakan algoritma *naïve bayes*. Penelitian ini akan mengidentifikasi dan memahami pandangan masyarakat mengenai *childfree* berdasarkan cuitan-cuitan di *twitter*. Penelitian ini menggunakan metode analisis sentimen dengan memanfaatkan media sosial *twitter*. Data diambil melalui proses *crawling* pada periode 19 desember 2022 hingga 16 januari 2023. Data yang diper*oleh* diolah menggunakan teknik *term frequency-inverse document frequency* (tf-idf) untuk menghasilkan 974 data yang siap untuk dianalisis. Algoritma *naïve bayes* digunakan untuk klasifikasi sentimen, dengan variasi data *training* dan data *testing* pada simulasi 60:40, 70:30, 80:20, dan 90:10. Mayoritas sentimen yang muncul adalah positif (mendukung *childfree*) dengan persentase tertinggi pada simulasi 4 (90:10) mencapai 90,72%. Sentimen positif ini mencerminkan dukungan terhadap kebebasan individu, pertimbangan finansial, kesejahteraan mental, serta pengakuan terhadap peran keluarga dalam keputusan *childfree*. Hasil penelitian ini dapat memberikan pemahaman lebih lanjut tentang faktor-faktor yang mempengaruhi keputusan *childfree*, serta dampaknya dalam berbagai konteks, termasuk hubungan individu, kebijakan publik, dan dinamika sosial.**

***Kata kunci*— analisis sentimen, *childfree*, *rapidminer*, *naïve bayes*, *text processing*.**

I. PENDAHULUAN

Kepopuleran sosial media *Twitter* di Indonesia tidak usah diragukan lagi, karena 59% penduduk Indonesia menggunakan *Twitter*, bahkan jumlah pengguna *Twitter* di Indonesia menduduki peringkat ke-5 sebagai media sosial yang paling banyak digunakan pada tahun 2020. *Twitter* sendiri dibangun oleh Jack Dorsey yang fungsi utamanya adalah mengirim dan menerima pesan maupun pendapat pengguna dengan istilah kicauan atau Tweet [1].

Twitter tidak terbatas waktu dan ruang, bahkan mampu menerima dan mengirim informasi dengan sangat cepat. Salah satu topik yang sempat ramai diperbincangkan pengguna *Twitter* di Indonesia adalah mengenai *childfree* setelah adanya berita mengenai Nomin seorang selegram asal korea selatan yang telah memilih *childfree* selama 7 tahun namun akhirnya memiliki anak karena ketidaksengajaan dan juga pengakuan Gita Savitri mengenai keinginannya untuk *childfree* [2].

Berdasarkan fenomena mengenai *childfree* tersebut, dalam penelitian ini memanfaatkan media sosial *Twitter* dalam menganalisis sentimen dari masyarakat yang mengeluarkan pendapatnya mengenai *childfree* [3]. Analisis sentimen ini dengan menggunakan metode *Text Mining* yang bertujuan untuk memproses data yang tidak terstruktur sehingga menghasilkan informasi yang dibutuhkan [4]. Setelah melakukan *Text Mining*, pada penelitian ini menganalisis sentimen menggunakan algoritma *Naïve Bayes* mengenai fenomena dari cuitan cuitan pengguna *Twitter* [5].

II. KAJIAN TEORI

A. *Text Mining*

Text Mining adalah sebuah langkah langkah untuk mencari pengetahuan yang berfokus pada data teks maupun dokumen dan bertujuan untuk menyaring informasi yang berguna untuk menemukan informasi penting dari informasi yang tidak terstruktur [6].*Text Mining* juga bisa disebut dengan pencarian informasi atau pengetahuan berupa data yang berbentuk teks dengan menghubungkan ketertarikan pada pengetahuan yang baru dibuat maupun baru terjadi [7].

B. *Data Mining*

Data Mining adalah sebuah langkah langkah menemukan kecocokan baru yang memiliki makna, pola, dan tren dengan menyaring sejumlah besar data yang tersimpan di dalam repositori, dengan menggunakan teknis teknik statistika serta pola penalaran yang ada pada teknologi [8].

C. *Text Preprocessing*

Text Preprocessing adalah tahapan pada pemrosesan data agar menjadi data yang siap dianalisis, dan menjadikan data lebih terstruktur [9]. Berikut merupakan tahapan text preprocessing :

1. *Case Folding* adalah tahapan mengubah seluruh teks tang berhuruf kapital menjadi huruf kecil.
2. *Tokenzing* adalah proses yang dilakukan untuk menghilangkan angka, tanda baca, dan karakter lain yang tidak memiliki pengaruh pada pemrosesan teks.
3. *Stopword Removal* adalah proses pemilihan kata yang dianggap penting dengan karakteristik kata yang mempunyai frekuensi kemunculan yang tinggi, seperti kata penghubung. Dan *Stopword Removal* bertujuan untuk mengurangi jumlah kata pada dokumen.
4. *Stemming* adalah proses mengubah kata yang ada pada sebuah dokumen ke kata kata akarnya *(root word)* dengan menggunakan aturan tertentu [9].

D. *Term Frequency-Inverse Document Frequency* (TF-IDF

Term Frequency-Inverse Document adalah sebuah perhitungan secara statistik yang digunakan untuk mengetahui seberapa pentingnya sebuah kata pada suatu dokumen, dan dokumen tersebut berada pada kelompok dokumen [10]. Pembobotan TF-IDF ini kemudian digunakan pada *text mining* [11]. Berikut adalah Rumus TF-IDF :

$$tf(t,d) = 0.5 + \frac{0{,}5\ x\ f(t,d)}{Maximum\ Occurences\ of\ word} \quad (2.1)$$

Dengan :
tf (t,d) = *term frecuency* kata t pada dokumen d
f(t,d) = jumlah frekuensi kata t pada dokumen d

$$idf(t,d) = log\frac{|D|}{no\ of\ documents\ term\ t\ appears} \quad (2.2)$$

Dengan :
Idf(t,f) = *Inverse Document frecuency* kata t dalam dokumen d
|D| = jumlah dokumen

E. Algoritma *Naïve Bayes*

Algoritma *Naive Bayes* adalah Algoritma klasifikasi yang melakukan pendekatan dengan mengacu pada teorema bayes, dengan mengolaborasikan pengetahuan sebelumnya dengan pengetahuan yang baru, sehingga memiliki akurasi yang tinggi [12].

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)} \quad (2.3)$$

Keterangan:
P (B|A) = Probabilitas Posterior, Probabilitas A diketahui maka, Probabilitas B muncul
P (A|B) = Probabilitas Posterior, Probabilitas B diketahui maka, ProbabilitasA muncul
P(A) = Probabilitas prior, probabilitas kejadian A
P(B) = Probabilitas prior, Probabilitas B [13]

F. *RapidMiner*

RapidMiner adalah sebuah perangkat lunak yang berguna pada pengolahan *Data Mining* untuk melakukan pengujian terhadap keakuratan klasifikasi mengenai data yang sedang dianalisis, *RapidMiner* juga mendukung 22 format file untuk dianalisis, seperti .xls, .csv, dan lain lain [14].

G. *Natural Language Processing*(NLP)

Natural Language Processing (NLP) adalah salah satu bidang komputer yang berfokus pada bidang kecerdasan buatan, dan bahasa (linguistik) diantaranya berhubungan dengan interaksi antara komputer dan bahasa alami manusia, seperti bahasa Inggris maupun bahasa Indonesia, Natural *Language Processing* memiliki tujuan yakni membuat mesin mengerti dan memahami arti bahasa manusia kemudian memberikan respon yang sesuai [15].

H. Evaluasi Performansi

Evaluasi Performansi adalah proses yang digunakan untuk menguji hasil dari sebuah klasifikasi dengan cara mengukur nilai performansi yang berasal dari sistem yang telah dibuat. Salah satu yang digunakan dalam evaluasi Performansi *data mining* adalah *Confusion Matrix* [16].

TABEL 1.1
Tabel *confusion matrix* klasifikasi tiga kelas

		Kelas Sebenarnya		
		Positif	Netral	Negatif
Kelas Prediksi	Positif	a	b	c
	Netral	d	e	f
	Negatif	g	h	i

Keterangan :

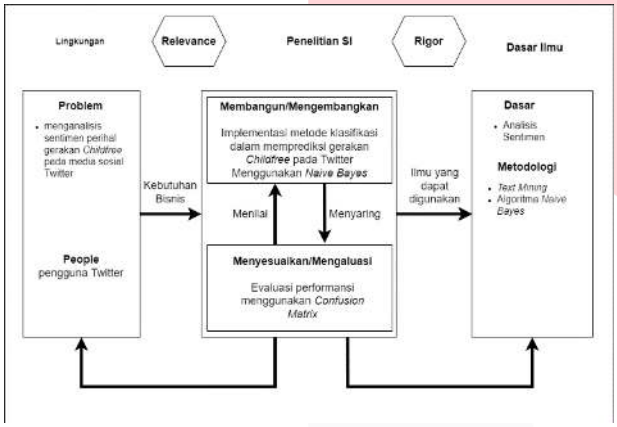
1. *True* Positif, apabila nilai sebenarnya positif dan hasil dari prediksi adalah positif.
2. Apabila nilai sebenarnya netral, tetapi hasil prediksi positif.
3. Apabila nilai sebenarnya negatif, tetapi hasil prediksi positif.
4. Apabila nilai sebenarnya positif, tetapi hasil prediksi netral.
5. *True* Netral, apabila nilai sebenarnya netral dan hasil dari prediksi adalah netral.
6. Apabila nilai sebenarnya adalah negatif dan hasil prediksi netral.
7. Apabila nilai sebenarnya positif, tetapi hasil prediksi negatif.
8. Apabila nilai sebenarnya netral, tetapi hasil prediksi negatif.

9. *True* Negatif, apabila nilai sebenarnya negatif dan hasil dari prediksi adalah negatif [17].

III. METODE PENELITIAN

A. Model konseptual

Pada penelitian ini dilakukan untuk menghasilkan informasi tentang fenomena *Childfree* di Indonesia. Proses penelitian ini diawali dengan pengidentifikasian masalah, kemudian menentukan rumusan masalah, kemudian dilakukan pengambilan data dari sosial media *Twitter*, dan perolehan data dengan menggunakan kata kunci yang diinginkan, lalu menganalisis data untuk melakukan deskripsi dari data yang telah didapat agar mudah dimengerti

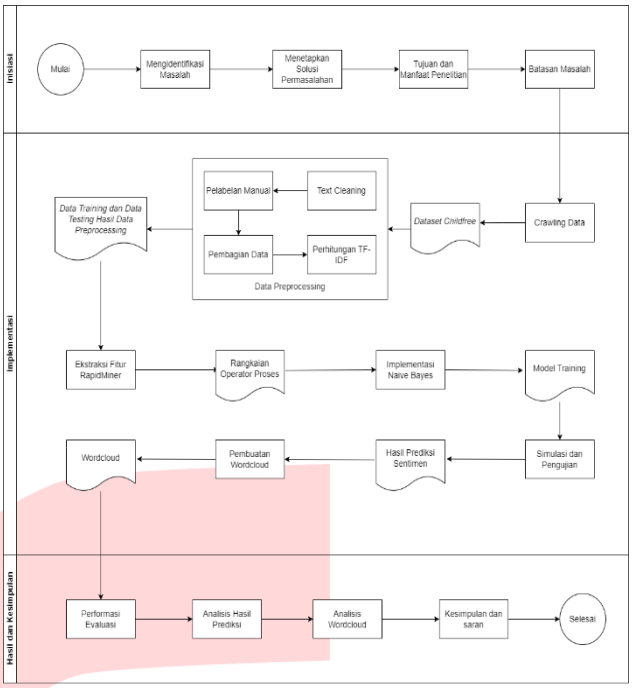


GAMBAR 2.1
Model Konseptual

Pada Gambar 1 diatas, Dijelaskan bahwa pada bagian lingkungan aspek people yakni pengguna *Twitter* yang dimaksud adalah para pengguna media sosial media *Twitter* yang memberikan komentar maupun tweet yang berkaitan dengan *chidfree*. Dan kemudian pada bagian problem adalah menganalisis sentimeni perihal fenomena *childfree* media sosial *Twitter*.

B. Sistematika Penyelesian Masalah

Pada penelitian ini terdapat beberapa tahapan mengenai sistematika penelitian, yakni inisiasi, Implementasi, dan hasil kesimpulan. Dibawah ini adalah penjelasan mengenai Sistematika Penelitian yang terdiri dari Inisiasi.



GAMBAR 2.2
Sistematika Penyelesaian Masalah

1. Inisiasi

Pada Proses hal pertama yang dilakukan adalah mengidentifikasi masalah mengenai *Childfree* yang dibahas oleh para pengguna media sosial *Twitter*. Kemudian menetapkan masalah akan digunakan dalam identifikasi masalah pada penelitian ini berdasarkan hasil dari studi literatur. Sehingga dapat menemukan tujuan dan manfaat dari penelitian ini,

2. Implementasi

Tahapan Implementasi pada penelitian ini diawali dengan mengumpulkan data yang berasal dari komentar maupun tweets para pengguna media sosial twitter yang berhubungan dengan topik pada penelitian ini yakni mengenai Fenomena *Childfree* dan beberapa hashtag yang berkaitan dengan *Childfree*.

3. Hasil dan Kesimpulan

Pada Bagian ini menggunakan tabel confusion matrix untuk melakukan proses performansi dengan pemetaan hasil prediksi sehingga memudahkan tahapan perhitungan yakni perhitungan precision yakni membandingkan jumlah prediksi benar dengan total prediksi. Dan akhir dari tahapan evaluasi adalah menganalisis hasil evaluasi performansi sehingga mendapatkan kesimpulan dan saran dari penelitian yang telah dibuat.

IV. PENGOLAHAN DATA DAN INPLEMENTASI

A. Crawling Data

Proses pengumpulan data dilakukan dengan menyaring opini-opini pada *Twitter* dalam bentuk (*tweet*) yang dipublikasi oleh pengguna *Twitter* di media sosial *Twitter* dengan topik pembahasan mengenai *childfree.* Pada penelitian ini crawling data dilakukan menggunakan *RapidMiner* dengan process seperti pada Gambar 3.1

GAMBAR 3. 1
Crawling process pada *RapidMiner*

Proses *crawling* dilakukan selama 29 hari yaitu mulai tanggal 19 Desember 2022 hingga 16 Januari 2023 dan dihasilkan total 1.474 data opini/*tweet* mengenai *childfree* dalam bahasa Indonesia yang kemudian digunakan untuk analisis sentimen.

B. Data Selection

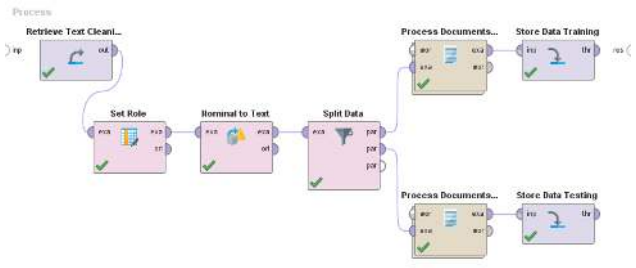
Dataset hasil *crawling* yang dihasilkan terdiri dari 12 atribut seperti tanggal pembuatan, nama *user*, hingga ID *tweet* itu sendiri. Pada dasarnya hanya data opini saja yang diperlukan dalam melakukan analisis sentimen. Pada penelitian ini hanya atribut "*text*" yang diperlukan karena berisi opini masyarakat mengenai fenomena *childfree* itu sendiri.

TABEL 3.1
Dataset *selection* yang digunakan.

No	*Text*
1	*Childfree* deh, demi kenyamanan acchi https://t.co/zbgZRthl2b
2	RT @BuahBuku: Sebagian orang ingin menunda punya anak, sebagian kecil lainnya benar-benar tidak ingin punya anak.
3	Habis nonton Teman Tapi Menikah 2. Film yank yankan yang menyenangkan, cerita waktu lahiran benar benar membuat saya heran, kok ada ya orangk yang masih mikir mau *childfree*.
4	RT @0kidokiy: @convomfs *no problem*, besarin anak itu susah, hrs siap mental, fisik dan materi.
…	…
1.474	Ternyata emak gw pengen cucu dari gw, sedangkan gw pengennya *childfree* lagian calonnya jg blm ada. Ini kudu ngomong gimana ma emak gw ya biar beliau ngerti ??

C. Data Prerocessing

Pada tahap ini, data hasil *crawling* yang telah mengalami seleksi dan diolah melalui serangkaian proses. Tahapan-tahapan tersebut mencakup *text cleaning*, konversi dari format nominal ke teks, serta pengolahan dokumen menggunakan metode *Term Frequency-Inverse Document Frequency* (TF-IDF).



GAMBAR 3. 2
Skema tahap akhir data *preprocessing* pada *RapidMiner*

D. Pemberian Label Manual (sentimen)

Dalam analisis sentimen mengenai *childfree* terdapat 3 kategori sentimen yaitu positif (mendukung *childfree*), netral (tidak memihak), atau negatif (menolak *childfree*). Dan menunjukkan gambaran umum pandangan masyarakat Indonesia mengenai fenomena *childfree*. Hasil pelabelan manual dijadikan acuan untuk algoritma mempelajari pola (data *training*) dan menghitung performa (data *testing*).

Tabel 3.2 Dataset *selection* yang digunakan.

No	*Text*	Sentimen
1	*Childfree* deh demi kenyamanan acchi	positif
2	Sebagian orang ingin menunda punya anak sebagian kecil lainnya benarbenar tidak ingin punya anak Kelompok kedua biasa dise…	netral
3	Habis nonton Teman Tapi Menikah 2 Film yank yankan yang menyenangkan cerita waktu lahiran benar benar membuat saya heran kok ada ya orangk yang masih mikir mau *childfree*	negatif
4	no problem besarin anak itu susah hrs siap mental fisik dan materi bener kata mamanya dara di film dua garis bi…	positif
…	…	
974	Ternyata emak gw pengen cucu dari gw sedangkan gw pengennya *childfree* lagian calonnya jg blm ada Ini kudu ngomong gimana ma emak gw ya biar beliau ngerti	positif

E. Pembuatan Data Testing dan Data Training

Pada penelitian ini dilakukan 4 simulasi pengujian dengan memvariasikan jumlah *data training* dan *data testing*. Simulasi ini dilakukan untuk mengetahui pengaruh jumlah persentase *data training* pada performa klasifikasi sehingga dapat menghasilkan akurasi terbaik. Pembagian *data training* dan *data testing* dalam berbagai variasi simulasi ditunjukkan pada Tabel dibawah ini :

TABEL 3. 3
Variasi *data training* dan *data testing* pada setiap simulasi.

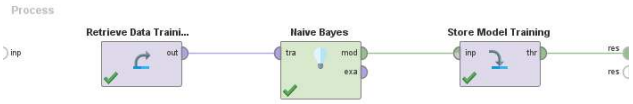
Simulasi	Data *Training*		Data *Testing*	
	Persentase	Jumlah Data	Persentase	Jumlah Data
1	60%	584	40%	390
2	70%	682	30%	292
3	80%	779	20%	195
4	90%	877	10%	97

F. Proses Analisis Sentimen

Performa klasifikasi dengan *Naive Bayes* diukur menggunakan *confusion matrix* dengan membandingkan hasil pemberian sentimen dari kondisi aktual (berdasarkan pelabelan manual) dan kondisi prediktif (berdasarkan hasil klasifikasi). Pembandingan kondisi aktual dan kondisi prediktif tersebut dilakukan pada data *training* dikarenakan data *training* teleh diberi label secara manual sehingga dapat dibandingkan dengan data prediktif.

G. Pembuatan Model Naïve Bayes

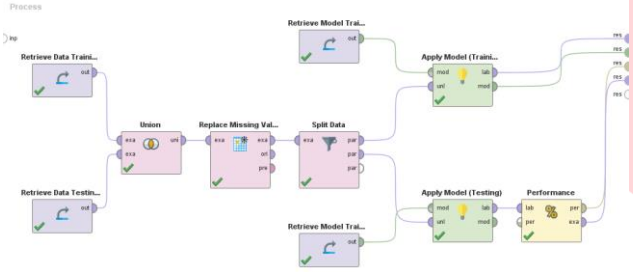
Tahap ini dilakukan untuk menghasilkan *model training* yang kemudian digunakan pada tahap simulasi. Pada tahap ini data *training* hasil data *preprocessing* yang telah diberi sentimen secara manual dipelajari dengan algoritma *Naïve Bayes*.



GAMBAR 3.3
Process pembuatan model dan data *training* pada *RapidMiner*.

H. Proses Simulasi Sentimen

Proses simulasi analisis sentimen "*childfree*" dilakukan dengan klasifikasi data *testing* menggunakan *machine learning* algoritma *Naive Bayes* serta perhitungan performa menggunakan *confusion matrix* pada data *testing*.



GAMBAR 3.4
Process simulasi pada analisis sentimen.

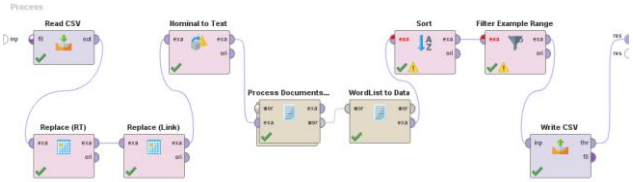
Prediksi dari pelabelan dihasilkan dari proses *machine learning* dengan algoritma *Naive Bayes*, dimana data *training* digunakan untuk melatih algoritma dalam mengklasifikasikan data selanjutnya menggunakan perhitungan TF-IDF.

Row No.	Sentimen	prediction(S...	confidence(...	confidence(...	confidence(...	text	aamin	aamiintapi	abdul
1	negatif	negatif	0	0	1	childfree menunda fikirkan anak merasakan hidup s...	0	0	0
2	positif	positif	1	0	0	pngn childfree bayi butuh biaya bnyk imunisasinya m...	0	0	0
3	netral	netral	0	1	0	wil childfree alasan pertimbangan financial mental a...	0	0	0
4	netral	netral	0	1	0	pribadi condong childfree dipikir kebanyakan diskusi ...	0	0	0
5	netral	netral	0	1	0	overpopulasi salah childfree beban realita kedepann...	0	0	0
6	positif	positif	1	0	0	childfree takut orang anak takut anakku ayah tanggun...	0	0	0
7	positif	positif	1	0	0	sadar anak mudah anak butuh dididik dirawat diseko...	0	0	0
8	netral	netral	0	1	0	gatau childfree engga ngomong temen suka bilang k...	0	0	0
9	positif	positif	1	0	0	jujur childfree ortu nuntut cucu	0	0	0
10	netral	netral	0	1	0	childfree gatau faktor	0	0	0
11	positif	positif	1	0	0	semoga cewek nerima anak childfree ekosistem bu...	0	0	0
12	positif	positif	1	0	0	childfree penyakit kelainan alasan tindakan perselus...	0	0	0
13	negatif	negatif	0	0	1	childfree nunda mulutnya dijaga	0	0	0
14	netral	netral	0	1	0	childfree nikah ewita suka anak orangnya int petit pny...	0	0	0

Gambar 3.5 Tampilan *ExampleSet* hasil klasifikasi *Naïve Bayes* untuk data *testing* pada *RapidMiner*.

I. Pembuatan Wordcloud

Dalam analisis sentimen, *wordcloud* diperlukan untuk menganalisis kata-kata mana saja yang sering muncul dalam pembahasan topik sentimen itu sendiri. Proses pembuatan *wordcloud* pada *RapidMiner* untuk sentimen mengenai "*childfree*" ditunjukkan pada Gambar

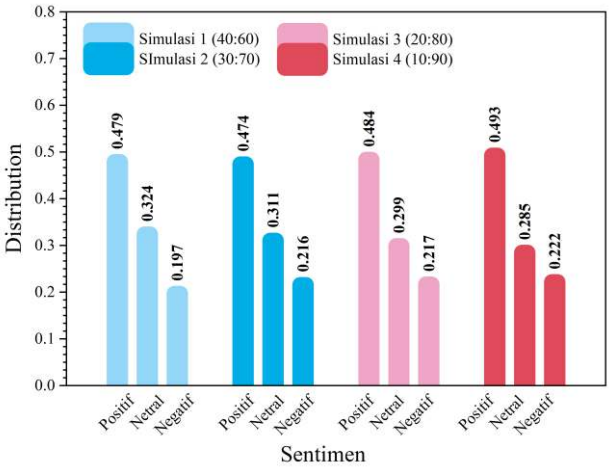


Gambar 3.6
Process pembuatan *wordcloud* pada *RapidMiner*.

V. HASIL DAN PEMBAHASAN

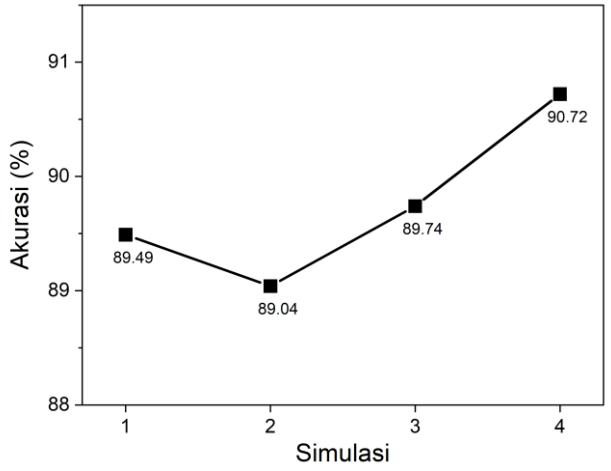
A. Analisis Hasil Simulasi Sentimen

Untuk mengetahui statistik sentimen yang paling sesuai dengan kondisi aktual mengenai opini masyarakat Indonesia terkait fenomena *childfree*, hasil simulasi 1, 2, 3, dan 4 akan memberikan gambaran tentang distribusi sentimen yang paling sesuai dengan kondisi aktual, sehingga dapat diambil kesimpulan mengenai opini masyarakat terhadap fenomena *childfree* berdasarkan hasil simulasi yang dilakukan.



GAMBAR 4.1
Perbandingan distribusi atribut pada kelas sentimen pada simulasi 1, 2, 3, dan 4.

Dalam perbandingan ini, terlihat bahwa semakin sedikit jumlah data *training* yang digunakan (mulai dari simulasi 2 hingga 4), tingkat akurasi yang dihasilkan semakin tinggi. Akurasi tertinggi terjadi pada simulasi 4. Semakin sedikit data *training* yang digunakan, algoritma menjadi lebih mudah dalam mengidentifikasi pola dan menghasilkan prediksi yang lebih akurat, yang mengakibatkan tingkat *error* yang lebih rendah.



Gambar 4.2
Kurva perbandingan tingkat akurasi.

B. WordCloud

Wordcloud adalah visualisasi yang menampilkan kata-kata dari teks yang diberikan dengan ukuran *font* yang lebih besar untuk kata-kata yang lebih sering muncul dalam teks. Dalam memahami sentimen mengenai *childfree*, diperlukan kata-kata yang menjadi kata kunci mengenai opini *childfree*

tersebut. Deretan kata tersebut disebut sebagai *wordlist* yang dihasilkan melalui proses TF-IDF. Terdapat 3.582 data *wordlist* yang dihasilkan dari data awal hasil *crawling* (1.474 data). Dipilih 30 data dengan frekuensi terbanyak yang kemudian digunakan sebagai data pembuatan *wordcloud*. *Wordcloud* yang dihasilkan ditunjukkan pada Gambar 3.5 dibawah ini :

GAMBAR 4.3
Wordcloud opini *childfree* pada media sosial twitter.

Berdasarkan analisis sentimen yang dilakukan, mayoritas masyarakat bersentimen positif terhadap *childfree*, dengan dukungan yang tampak dalam kata-kata seperti "setuju", "pasangan", dan "sepasang". Sebagian masyarakat bersentimen netral, terlihat dari penggunaan kata-kata seperti "takut" dan "alasan" yang mencerminkan pemahaman terhadap keputusan individu. Minoritas masyarakat bersentimen negatif, yang tercermin dalam kata-kata seperti "buang".

VI. KESIMPULAN DAN SARAN

A. Kesimpulan

Proses analisis sentimen dilakukan dengan *crawling* data terlebih dahulu pada media sosial *Twitter* menggunakan *RapidMiner*. Dihasilkan 1.474 data yang diper*oleh* mulai dari tanggal 19 Desember 2022 hingga 16 Januari 2023. Data yang dihasilkan diolah dengan *text processing* menggunakan *Term Frequency-Inverse Document Frequency* (TF-IDF) pada 4 variasi data *training* dan data *testing* dengan komposisi 60:40 pada simulasi 1, 70:30 pada simulasi 2, 80:20 pada simulasi 3, dan 90:10 pada simulasi 4.

Hasil analisis sentimen pada 4 variasi data *training* dan data *testing* menunjukkan sentimen mayoritas yang dihasilkan bersifat positif (410 data dari 974) yaitu mendukung adanya *childfree* dengan alasan kebebasan individu, pertimbangan finansial, kesejahteraan mental, serta pengakuan terhadap peran keluarga.

Tingkat akurasi tertinggi yang dihasilkan berdasarkan hasil klasifikasi menggunakan algoritma *Naive Bayes* adalah 90,72% yang diper*oleh* dari hasil simulasi 4 dengan 90% data *training* dan 10% data *testing*. Hal tersebut menunjukkan keberhasilan klasifisikasi dengan tingkat akurasi yang tinggi.

B. Saran

Perluas data untuk mendapatkan representasi yang lebih komprehensif tentang pandangan masyarakat mengenai *childfree*. Lalu lakukan analisis kualitatif melalui wawancara atau studi kasus mendalam untuk memperkaya pemahaman tentang alasan, pengalaman, dan pandangan individu terhadap *childfree*.

Bandingkan pandangan masyarakat terhadap *childfree* dengan kelompok kontrol untuk memahami perubahan pola pikir dan penerimaan terhadap *childfree* dan Lakukan studi perbandingan dengan masyarakat di negara lain atau konteks budaya yang berbeda untuk mendapatkan perspektif lintas budaya yang berharga

REFERENSI

[1] T. Krisdiyanto, “Analisis Sentimen Opini Masyarakat Indonesia Terhadap Kebijakan PPKM pada Media Sosial Twitter Menggunakan Naïve Bayes Clasifiers,” *J. CoreIT J. Has. Penelit. Ilmu Komput. dan Teknol. Inf.*, vol. 7, no. 1, p. 32, 2021, doi: 10.24014/coreit.v7i1.12945.

[2] F. Fahusni, “Ramai Diperbincangkan, Childfree Jadi Trending Topic Artikel ini telah tayang di Selular.id Ramai Diperbincangkan, Childfree Jadi Trending Topic,” *selular.id*, 2023. https://selular.id/2023/02/ramai-diperbincangkan-childfree-jadi-trending-topic/.

[3] J. Moore, “Reconsidering Childfreedom: A Feminist Exploration of Discursive Identity Construction in Childfree LiveJournal Communities,” *Women’s Stud. Commun.*, vol. 37, no. 2, pp. 159–180, 2014, doi: 10.1080/07491409.2014.909375.

[4] R. Talib, M. Kashif, S. Ayesha, and F. Fatima, “Text Mining: Techniques, Applications and Issues,” *Int. J. Adv. Comput. Sci. Appl.*, vol. 7, no. 11, pp. 414–419, 2016, doi: 10.14569/ijacsa.2016.071153.

[5] H. Hassani, C. Beneki, S. Unger, M. T. Mazinani, and M. R. Yeganegi, “Text mining in big data analytics,” *Big Data Cogn. Comput.*, vol. 4, no. 1, pp. 1–34, 2020, doi: 10.3390/bdcc4010001.

[6] A. Ahadi, A. Singh, M. Bower, and M. Garrett, “Text Mining in Education—A Bibliometrics-Based Systematic Review,” *Educ. Sci.*, vol. 12, no. 3, 2022, doi: 10.3390/educsci12030210.

[7] D. Antons, E. Grünwald, P. Cichy, and T. O. Salge, “The application of text mining methods in innovation research: current state, evolution patterns, and development priorities,” *R D Manag.*, vol. 50, no. 3, pp. 329–351, 2020, doi: 10.1111/radm.12408.

[8] C. Romero and S. Ventura, “Educational data mining and learning analytics: An updated survey,” *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.*, vol. 10, no. 3, pp. 1–21, 2020, doi: 10.1002/widm.1355.

[9] V. Kalra and R. Aggarwal, “Importance of Text Data Preprocessing & Implementation in RapidMiner,” *Proc. First Int. Conf. Inf. Technol. Knowl. Manag.*, vol. 14, no. July, pp. 71–75, 2018, doi: 10.15439/2017km46.

[10] H. Patil and R. S. Thakur, “Document Clustering,” pp. 264–281, 2016, doi: 10.4018/978-1-5225-0536-5.ch013.

[11] I. Garcia-Magarino, G. Gray, R. Lacuesta, and J.

Lloret, "Survivability Strategies for Emerging Wireless Networks with Data Mining Techniques: A Case Study with NetLogo and RapidMiner," *IEEE Access*, vol. 6, no. 8, pp. 27958–27970, 2018, doi: 10.1109/ACCESS.2018.2825954.

[12] H. R. Pourghasemi, A. Gayen, S. Park, C. W. Lee, and S. Lee, "Assessment of landslide-prone areas and their zonation using logistic regression, LogitBoost, and naïvebayes machine-learning algorithms," *Sustain.*, vol. 10, no. 10, 2018, doi: 10.3390/su10103697.

[13] M. Faid, M. Jasri, and T. Rahmawati, "Perbandingan Kinerja Tool Data Mining Weka dan Rapidminer Dalam Algoritma Klasifikasi," *Teknika*, vol. 8, no. 1, pp. 11–16, 2019, doi: 10.34148/teknika.v8i1.95.

[14] Sudirman, A. P. Windarto, and A. Wanto, "Data mining tools | rapidminer: K-means method on clustering of rice crops by province as efforts to stabilize food crops in Indonesia," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 420, no. 1, 2018, doi: 10.1088/1757-899X/420/1/012089.

[15] M. Müller, M. Salathé, and P. E. Kummervold, "COVID-Twitter-BERT: A Natural Language Processing Model to Analyse COVID-19 Content on Twitter," 2020, [Online]. Available: http://arxiv.org/abs/2005.07503.

[16] R. Susmaga, "Confusion Matrix Visualization," in *Intelligent Information Processing and Web Mining*, Berlin, Heidelberg: Springer Berlin Heidelberg, 2004, pp. 107–116.

[17] C. Anam and H. B. Santoso, "Perbandingan Kinerja Algoritma C4.5 dan Naive Bayes untuk Klasifikasi Penerima Beasiswa," *Energy - J. Ilm. Ilmu-Ilmu Tek.*, vol. 8, no. 1, pp. 13–19, 2018, [Online]. Available: https://ejournal.upm.ac.id/index.php/energy/article/view/111.