

Erizal, Hasan - cek.01

by Cek Turnitin UHAMKA

Submission date: 01-Apr-2024 11:18AM (UTC+0700)

Submission ID: 2336604018

File name: JOSH_Erizal,_Hasan_-_cek.01.docx (497.78K)

Word count: 4343

Character count: 27695



Analisis Sentimen Masyarakat Terhadap Fenomena Childfree (Kehidupan Tanpa Anak) Pada Twitter Menggunakan Algoritma Naïve Bayes

Erizal¹, Firman Noor Hasan^{2*}

¹ Fakultas Teknologi Industri dan Informatika, Sistem dan Teknologi Informasi, Universitas Muhammadiyah Prof. Dr. Hamka, Jakarta, Indonesia

Jl. Tanah Merdeka, Pasar Rebo, 13710, Jakarta, Indonesia

² Fakultas Teknologi Industri dan Informatika, Teknik Informatika, Universitas Muhammadiyah Prof. Dr. Hamka, Jakarta, Indonesia

Jl. Tana Merdeka, Pasar Rebo, 13710, Indonesia

Email: ¹erizal@uhamka.ac.id, ²firman.noorhasan@uhamka.ac.id

Email Penulis Korespondensi: firman.noorhasan@uhamka.ac.id

Submitted: 99/99/999; Accepted: 99/99/999; Published: 99/99/999

Abstrak—*Childfree* merupakan salah satu fenomena yang terjadi tidak hanya di dunia maupun di Indonesia. Banyak stigma negatif maupun positif yang muncul terkait fenomena untuk menjalani kehidupan tanpa anak khususnya di perkotaan. Tanggapan masyarakat terhadap fenomena *childfree* khususnya pada perkotaan di Indonesia beragam, dan lebih banyak dipengaruhi oleh berbagai faktor. Pada penelitian ini, menganalisis pandangan netizen terhadap fenomena *childfree* menggunakan metode Naïve Bayes dibantu dengan tools rapidminer untuk mengolah data teks yang dikumpulkan melalui media social X. Tujuan dari penelitian ini adalah untuk menganalisis serta menyajikan data mengenai sentimen masyarakat terhadap fenomena *childfree* di Indonesia. Hasil dari penelitian mendapat sebanyak 319 yang merujuk kepada sentimen negatif, serta hanya sejumlah 181 yang merujuk kepada sentimen positif. Hasil akurasi yang dihasilkan oleh algoritma Naïve Bayes sebesar 95,02%. Memperlihatkan bahwa fenomena *childfree* dipilih oleh sebagian netizen khususnya diperkotaan karena mereka menginginkan masa muda untuk fokus kepada pendidikan, maupun karir demi membuat hidup mereka lebih nyaman.

Kata Kunci: Analisis Sentimen; Childfree; Twitter; Naïve Bayes;

Abstract—*Childfree* is a phenomenon that occurs not only in the world but also in Indonesia. There are many negative and positive stigmas that arise regarding the phenomenon of living a life without children, especially in urban areas. The public's response to the *childfree* phenomenon, especially in urban areas in Indonesia, is varied, and is more influenced by various factors. In this research, we analyzed netizens' views on the *childfree* phenomenon using the Naïve Bayes method assisted by rapidminer tools to process text data collected via social media X. The aim of this research is to analyze and present data regarding public sentiment towards the *childfree* phenomenon in Indonesia. The results of the research found that 319 referred to negative sentiment, and only 181 referred to positive sentiment. The accuracy results produced by the Naïve Bayes algorithm were 95.02%. Showing that the *childfree* phenomenon is chosen by some netizens, especially in urban areas, because they want young people to focus on education and careers in order to make their lives more stable.

Keywords: Sentiment Analysis; ChildFree; Twitter; Naïve Bayes;

1. PENDAHULUAN

Seiring dengan perkembangan teknologi, terutama pada jejaring sosial, Indonesia sedang disibukkan dengan fenomena baru yang hadir disela-sela kehidupan masyarakat. *Childfree* sendiri adalah bentuk pilihan hidup tanpa dikaruniai seorang anak atau dengan kata lain *childfree* merupakan pilihan hidup yang secara sukarela memilih untuk tidak memiliki anak dengan pertimbangan lingkungan [1]. Fenomena *childfree* yang datang dari kebudayaan luar ini perlahan hadir dan berkembang di Indonesia. Namun, mengingat beragamnya sudut pandang masyarakat dan banyaknya perbedaan lingkungan kehidupan sosial masyarakat, menjadikan fenomena ini masih banyak didebatkan kehadirannya. Ada yang menganggap fenomena ini cocok untuk diterapkan di Indonesia, dan begitupun sebaliknya, mengingat kembali bagaimana budaya yang diterapkan di sini. Beberapa masyarakat juga masih ada yang belum bisa menyaring kehadiran fenomena ini dengan baik. Hal ini ditunjukkan dengan adanya beberapa masyarakat yang sekadar mengikuti fenomena ini dikarenakan sedang menjadi trend [2].

Salah satu media sosial yang menjadi tempat banyak masyarakat untuk berbagi komentar atau memberikan opini mereka adalah Twitter. *Tweet* yang ditulis oleh masyarakat sangat berguna karena memberikan umpan balik terhadap suatu hal. Selain itu, *tweet-tweet* tersebut dapat dijadikan dasar dalam memahami sentimen atau persepsi masyarakat terhadap hal yang dibicarakan [3]. Hal ini dikarenakan Twitter memiliki lebih dari 140 juta pengguna aktif yang setiap harinya dapat membagikan lebih dari 400 juta *tweet* atau opini mereka kepada publik. Tak hanya itu, pengguna Twitter juga dapat membicarakan hal-hal yang sedang tren dan membagikannya untuk mendapatkan umpan balik dari pengguna lainnya [4].



Data-data *tweet* tersebut akan diproses bersama *text mining* untuk dilakukan analisis terhadapnya. *Text mining* sendiri bertujuan untuk mencari dan menemukan informasi yang belum diketahui sebelumnya, namun berpotensi berguna dari data teks tak terstruktur atau semi-struktur [5]. Dari berbagai macam *tweet* yang terdapat pada Twitter tersebut, tentunya perlu waktu yang lama untuk dilakukan proses analisis. Dengan ini, beberapa metode dapat digunakan dengan tujuan menyingkat waktu proses analisis sentimen terhadap data *tweet* tersebut. Salah satu metode yang dapat digunakan untuk menganalisis sentimen suatu hal tersebut adalah metode atau algoritma Naïve Bayes [6], [7]. Algoritma Naïve Bayes didasari oleh prinsip Bayes yang menjelaskan bahwa seluruh kegiatan memiliki kontribusi yang kepentingannya sama atau memiliki kebebasan terhadap pemilihan *class* tertentu. Algoritma ini adalah metode yang digunakan dalam proses *text mining* untuk memvisualisasi sentimen masyarakat [8].

Analisis sentimen adalah metode pengidentifikasi sebuah sentimen yang berbentuk data teks dan bagaimana sentimen tersebut bisa dikategorikan sebagai sentimen yang positif atau negatif [9]. Dengan kata lain, analisis sentimen juga memiliki istilah lain yakni, *opinion mining*. *Opinion mining* sendiri merupakan gabungan dari *text mining* dan *natural language processing*. Tujuan dari *text mining* yaitu sebagai penambah data teks yang asalnya dari *file* tertentu dan mencari kata-kata yang *file* tersebut wakilkkan. Dengan demikian, memungkinkan adanya analisis yang terhubung antar *file* tersebut [10]. Nantinya proses analisis tersebut dilakukan menggunakan *tool* Rapid Miner menggunakan algoritma Naïve Bayes. Pada penelitian ini, Rapid Miner digunakan sebagai penganalisis data dan juga sebagai mesin *data mining* [11]. Pada Rapid Miner, dilakukan beberapa tahapan untuk memenuhi penelitian ini. Diawali dengan menarik dan mengumpulkan data-data berbentuk opini dari Twitter. Setelah melakukan tahap *crawling data* tersebut, dilakukan beberapa tahap lainnya guna membersihkan data-data tersebut agar dapat dianalisis dengan baik. Tahap *cleansing* dan *filtering* tersebut mencakupi tokenisasi, *case folding*, *stopword removal*. Setelahnya dilakukan pelabelan, dan pengklasifikasian [12][13]. Setelah melalui tahap tersebut dilakukan penganalisan menggunakan Naïve Bayes yang memungkinkan adanya pengklasifikasian yang didasari oleh asumsi di mana setiap *attributes* mempunyai hubungan kondisional yang mandiri di setiap kelasnya. Dengan demikian, algoritma ini adalah metode pengklasifikasian yang sangat efektif serta menghasilkan hasil klasifikasi yang baik [14]. Metode ini terbukti efektif dalam melakukan pengklasifikasian serta kinerjanya lebih baik dibandingkan dengan metode lainnya. Naïve bayes dinilai lebih baik karena kecepatan dan kesederhanaannya dalam mengklasifikasi suatu data teks [15]. Keuntungan lainnya yang diberikan oleh algoritma Naïve Bayes yakni tidak diperlukannya data latih dalam jumlah yang begitu banyak, sehingga proses klasifikasi teks yang akan diprediksi dapat dilakukan dengan mudah dan cepat. Untuk memperhitungkan klasifikasi metode ini adalah dengan perhitungan probabilitas [16].

Penelitian serupa dilakukan oleh Pristiyono berjudul “Sentiment Analysis of COVID-19 Vaccine in Indonesia using Naïve Bayes Algorithm” [17] yang membahas mengenai analisis sentimen terhadap vaksin untuk COVID-19 di Indonesia dengan menjadikan algoritma Naïve Bayes sebagai pengklasifikasinya. Penelitian ini bertujuan untuk menilai sentimen masyarakat Indonesia melalui analisis jaringan sosial dari vaksin COVID-19 per Januari 2021. Hasil pengukuran sentimen penelitian ini menunjukkan bahwa terdapat lebih dari 56% *tweet* negatif, lebih dari 39% *tweet* positif, dan 1% *tweet* netral. Penelitian selanjutnya yang serupa adalah penelitian yang dilakukan oleh Charlyn Villavicencio yang berjudul “Twitter Sentiment Analysis towards COVID-19 Vaccines in the Philippines Using Naïve Bayes” [18] yang membahas tentang analisis sentimen dengan algoritma Naïve Bayes mengenai vaksin COVID-19 di Filipina dengan polaritas positif, negatif, dan netral. Hasil dari penelitian ini menunjukkan bahwa algoritma Naïve Bayes menghasilkan nilai akurasi sebesar 81,77%. Penelitian lainnya yang juga serupa dilakukan oleh Lopamudra Dey yang berjudul “Sentiment Analysis of Review Datasets using Naïve Bayes and K-NN Classifier” [19] yang membahas mengevaluasi kinerja klasifikasi sentimen dalam hal akurasi, dan presisi. Algoritma yang digunakan untuk menganalisis topik dari penelitian itu adalah Naïve Bayes dan K-Nearest Neighbour. Hasil percobaan pada penelitian ini menunjukkan bahwa hasil dengan pendekatan Naïve Bayes lebih baik dari pendekatan K-NN dengan menghasilkan nilai akurasi sebesar 80%. Perbedaan penelitian ini dengan penelitian terdahulu adalah pengukuran sentimen, yang di mana pada penelitian terdahulu mengukur sentimen positif, negatif, dan netral. Sedangkan penelitian ini hanya mengukur dua jenis sentimen saja, yakni positif dan negatif. Selain itu, terdapat perbedaan pada penggunaan algoritma pengklasifikasi. Pada penelitian terdahulu, terdapat perbandingan antara algoritma Naïve Bayes dan K-Nearest Neighbour. Sedangkan penelitian ini hanya menggunakan algoritma Naïve Bayes sebagai penganalisis sentimen.

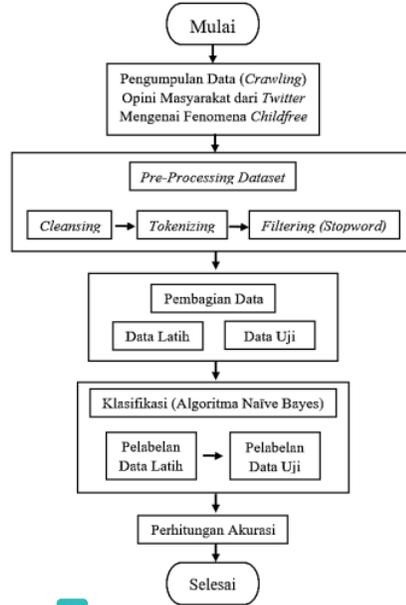
Pada penelitian ini, data-data yang dikumpulkan berasal dari media Twitter berisikan komentar atau opini-opini masyarakat mengenai *childfree*. Setelah itu, data yang telah dikumpulkan tersebut akan dilakukan proses pembersihan data, lalu setelahnya akan dianalisis dengan menggunakan algoritma Naïve Bayes. Penelitian ini sendiri memiliki tujuan, yakni mencari tahu sentimen positif dan negatif masyarakat Indonesia mengenai fenomena *childfree* menggunakan *tweet* yang diunggah melalui Twitter. Kemudian data-data yang telah teridentifikasi apakah positif atau negatif akan dilakukan analisis dengan menerapkan algoritma Naïve Bayes [20]. Penelitian ini fokus kepada opini masyarakat mengenai *childfree* yang terdapat pada Twitter berjumlah 500 data. Proses penarikan data dan analisis sentimen dilakukan menggunakan Rapid Miner.

13

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

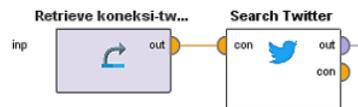
Tahapan dari penelitian ini diawali dari tahap data (*crawling*) dan dilanjutkan ke tahap analisis dari data yang telah melalui tahap-tahap sebelumnya.



16
Gambar 1. Diagram alir penelitian

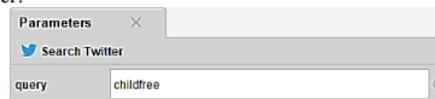
2.2 Pengumpulan Data

Tahap ini merupakan langkah awal yakni, tahap pengumpulan data berbentuk opini dari *Twitter* melalui Rapid Miner menggunakan operator *Search Twitter*. Data yang dikumpulkan untuk penelitian ini sejumlah 2000 data dengan rentang waktu pengumpulan data adalah pada tanggal 3 s.d. 10 Mei 2023.



Gambar 2. Crawling Data

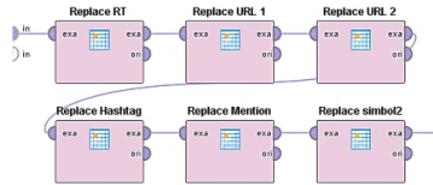
Pada gambar 2 terdapat proses pengumpulan data melalui *Rapid Miner* yang menggunakan dua operator utama yakni, operator *Retrieve* (pengambilan) dan operator *Search* (Pencarian). Dikarenakan penelitian ini ingin mengumpulkan data dari *Twitter*, maka digunakan operator *Retrieve* koneksi-twitter untuk melakukan koneksi dengan *twitter* agar dapat mengumpulkan data-data yang ingin digunakan untuk penelitian ini. Sedangkan operator *Search Twitter* adalah operator yang digunakan untuk mencari *tweet-tweet* ³⁷ *Twitter* mengikuti kata kunci, waktu, dan lainnya yang dimasukkan ke dalam parameter dari operator ini. Pada penelitian ini, penulis melakukan pengumpulan data menggunakan operator *Search Twitter* dan berdasarkan kata kunci “*childfree*” yang dimasukkan ke dalam parameter.



Gambar 3. Kata kunci pada operator untuk search twitter

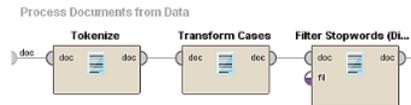
30
2.3 Pre-Processing

Setelah data terkumpul, dilakukan tahap *pre-processing* yang mencakup *cleansing data*, *tokenizing*, dan *filtering* (*stopword removal*). Tahap ini dilakukan dengan tujuan supaya proses klasifikasi atau analisis dapat diproses secara mudah [21]. Di bawah ini merupakan operator yang digunakan untuk melakukan pembersihan data tersebut. Operator ini adalah *Replace* yang fungsinya adalah untuk membersihkan data-data tersebut dari beberapa simbol, angka, link, dan lain-lain yang tidak terkait. Dalam proses pembersihan data pada penelitian ini, operator *Replace* digunakan untuk menghapus kata "RT", tautan-tautan, *mention*, simbol *hashtag*, dan simbol-simbol lainnya.



Gambar 4. Cleansing Data

Tahap selanjutnya merupakan *tokenizing*, *case folding*, dan *filtering stopwords*. Operator *Tokenize* yang ada pada gambar 4 di bawah ini merupakan operator yang digunakan sebagai pemecah atau pemisah sebuah teks berbentuk paragraf, kalimat, atau kata yang berimbuhan menjadi sebuah kata dasar. Operator *Transform Case* adalah operator untuk mengubah semua huruf pada teks menjadi huruf kecil (*lower case*) atau huruf besar (*upper case*) [22]. Sedangkan operator *Filtering Stopword* adalah operator yang digunakan untuk menghilangkan kata-kata yang terdapat di dalam *stopword*. Kata-kata yang terkandung dalam *stopword* tersebut merupakan kata-kata yang tidak memiliki arti atau kontribusi penting dalam proses penganalisisan. Ketiga operator tersebut dimasukkan ke dalam operator bernama *Process Document from Data*. Operator ini dirancang untuk menyiapkan data-data teks yang nantinya akan dilakukan penganalisisan lebih lanjut menggunakan algoritma yang digunakan.



Gambar 5. Tokenize, Case folding, Stopword

Setelah melalui tahap *pre-processing*, dan menghilangkan beberapa data yang memiliki duplikasi, data yang tadinya berjumlah 2000 hanya tersisa menjadi 500 data teks berbentuk opini.

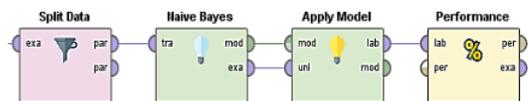
2.4 Pelabelan Data

Setelah *pre-processing*, pembagian data menjadi 2 bagian, yakni *data training* dan *data testing* akan dilakukan. Setelahnya, *data training* atau data latih akan melalui tahap labelisasi secara independen, dan nantinya data tersebut digunakan sebagai acuan analisis bagi data testing atau uji yang akan dilakukan langsung di Rapid Miner.

2.5 Klasifikasi Data Uji (Naive Bayes)

Setelah *pre-processing*, pembagian data menjadi 2 bagian, yakni *data training* dan *data testing* akan dilakukan. Setelahnya, *data training* atau data latih akan melalui tahap labelisasi secara independen, dan nantinya data tersebut digunakan sebagai acuan analisis bagi data testing atau uji yang akan dilakukan langsung di Rapid Miner.

Setelah latih telah terlabeli semua, kemudian data tersebut akan digunakan sebagai acuan untuk prediksi sentimen dan pengujian terhadap data *testing* menggunakan algoritma Naive Bayes di dalam Rapid Miner.



Gambar 6. Klasifikasi Data

Da gambar 6 terdapat operator *Split Data*, *Naive Bayes*, *Apply Model*, dan *Performance*. Operator *Split Data* digunakan untuk membagi dataset menjadi 2 bagian yakni, data latih dan data uji. Pada penelitian ini saya

menggunakan rasio 60:40 dalam pembagian datanya, yakni 60% untuk data latih dan 40% data uji. Selanjutnya adalah algoritma pengklasifikasi Naïve Bayes yang digunakan untuk mengklasifikasikan atau menganalisis sentimen. Lalu, operator Apply Model digunakan sebagai machine learning untuk memprediksi data uji. Operator terakhir, yakni operator Performance yang merupakan operator yang melakukan penilaian terhadap kinerja model machine learning. Jadi, keempat operator tersebut merupakan operator-operator yang digunakan dalam proses persiapan data, pengembangan model, penerapan model pada data uji, dan pengevaluasi kinerja model dalam hal menganalisis data. Dari proses klasifikasi data tersebut didapatkan hasil prediksi sentimen terhadap data uji sebagai berikut:

Tabel 1. Jumlah Data

Jenis Data	Positif	Negatif
Data Latih	223	79
Data Uji	100	105
Total	323	184

2.4 Evaluasi Hasil Pengujian

Setelah proses pengujian dan prediksi sentimen, dapat dilakukan perhitungan probabilitas. Terdapat 2 perhitungan probabilitas yakni, probabilitas prior dan posterior. Untuk melakukan perhitungan probabilitas posterior, harus terlebih dulu dilakukan perhitungan probabilitas prior.

Perhitungan probabilitas prior ini dilakukan dari setiap class yang terdapat di dalam dataset yang digunakan pada penelitian ini. Tujuannya adalah untuk menghindari predisposisi dalam klasifikasi terhadap pengelompokan class yang umum ditemukan hingga yang jarang muncul. Probabilitas prior akan menghasilkan class positif dan negatif yang diperoleh dari proses identifikasi mengenai fenomena *childfree* [23].

- Probabilitas Prior Positif

$$P(\text{Positif}) = \frac{\text{Jumlah Data Positif}}{\text{Total Jumlah Data}} \quad (1)$$

- Probabilitas Prior Negatif

$$P(\text{Negatif}) = \frac{\text{Jumlah Data Negatif}}{\text{Total Jumlah Data}} \quad (2)$$

Setelah menghasilkan probabilitas prior, lalu dilakukan perhitungan probabilitas posterior. Perhitungan probabilitas posterior ini bertujuan untuk menentukan class terhadap hasil identifikasi kasus baru. Hal ini dilakukan jika teridentifikasi adanya kasus baru dalam pengolahan data pada penelitian ini [24].

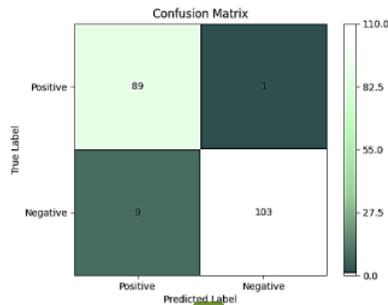
- Probabilitas Posterior Positif

$$\text{Posterior Pos} = \text{Class Positif} \times P(\text{Positif}) \quad (3)$$

- Probabilitas Posterior Negatif

$$\text{Posterior Neg} = \text{Class Negatif} \times P(\text{Negatif}) \quad (4)$$

Selain probabilitas, nilai *accuracy*, *precision*, dan *recall* juga kemudian dilakukan analisa. Nilai-nilai tersebut dapat di perhitungkan dengan terlebih dulu memperhitungkan *confusion matrix* yang tertera pada tabel dan gambar berikut:



Gambar 7. Confusion Matrix

Confusion matrix di atas merupakan tabel yang isinya berupa pernyataan dari klasifikasi jumlah data yang benar terprediksi positif atau True Positive (TP), salah terprediksi positif atau False Positive (FP), benar terprediksi



negatif atau *True Negative* (TN), dan salah terprediksi negatif atau *False Negative* (FN). Nantinya, *confusion matrix* tersebut akan digunakan untuk menghitung nilai *accuracy*, *recall* (*True Positive Rate*), *precision* (*Positive Predictive Value*), *True Negative Rate*, dan *Negative Predictive Value*.

Nilai *accuracy* merupakan persentase dari data uji atau data *testing* yang kelasnya berhasil diidentifikasi atau diklasifikasikan dengan benar oleh sistem berdasarkan kelas aslinya. Rumus perhitungan nilai *accuracy* dapat dilihat pada persamaan (5) di bawah ini [25]:

$$Accuracy = \frac{TP+TN}{TP+FN+FP+TN} \quad (5)$$

Nilai *recall* atau nilai *True Positive Rate* yang memiliki nama lain, yakni nilai *sensitivity* merupakan persentase dari keberhasilan sistem dalam mengklasifikasi kelas positif sebagai kelas positif. Rumus perhitungan nilai *recall* (TPR) terdapat pada persamaan (6) sebagai berikut [25]:

$$Recall (True Positive Rate) = \frac{TP}{TP+FN} \quad (6)$$

Nilai *precision* atau nilai *Positive Predictive Value* merupakan persentase dari data yang terprediksi sebagai kelas positif oleh algoritma pengklasifikasi yang merupakan data positif sebenarnya dari seluruh yang terprediksi sebagai kelas positif. Rumus perhitungan nilai *precision* (PPV) terdapat pada persamaan (7) sebagai berikut [25]:

$$Precision (Positive Predictive Value) = \frac{TP}{TP+FP} \quad (7)$$

Nilai *True Negative Rate* yang memiliki nama lain, yakni nilai *specificity* merupakan persentase dari keberhasilan sistem dalam mengklasifikasi kelas negatif sebagai kelas negatif. Rumus perhitungan nilai TNR terdapat pada persamaan (8) sebagai berikut [26].

$$True Negative Rate = \frac{TN}{FP+TP} \quad (8)$$

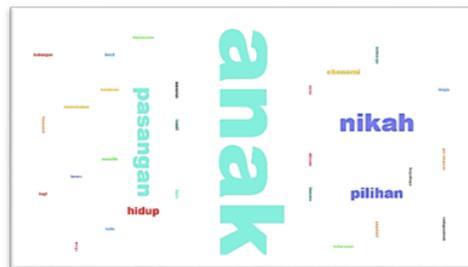
Nilai NPV atau nilai *Negative Predictive Value* merupakan persentase dari data yang terprediksi sebagai kelas negatif oleh algoritma pengklasifikasi yang merupakan data negatif sebenarnya dari seluruh yang terprediksi sebagai kelas negatif. Rumus perhitungan nilai NPV atau nilai *Negative Predictive Value* terdapat pada persamaan (9) sebagai berikut [26]:

$$Negative Predictive Value = \frac{TN}{FN+TN} \quad (9)$$

3. HASIL DAN PEMBAHASAN

Pada bagian ini, hasil dari penelitian dan beberapa pengujian dilakukan. Data-data yang telah teranalisis sentimennya menggunakan algoritma Naïve Bayes, kemudian dilakukan perhitungan probabilitasnya. Perhitungan probabilitas yang digunakan pada penelitian ini adalah probabilitas prior dan probabilitas posterior. Setelah melakukan perhitungan terhadap nilai probabilitas prior dan posterior, dilakukan pula perhitungan terhadap nilai *accuracy*, *recall* (TPR), *precision* (PPV), TNR, dan NPV dengan menjadikan *confusion matrix* sebagai acuan dalam melakukan kelima perhitungan setelah perhitungan probabilitas.

Untuk memperhitungkan kedua nilai probabilitas yakni, probabilitas prior dan posterior dibutuhkan kata-kata yang paling sering muncul dari data yang dikumpulkan. Berikut adalah visualisasi *wordcloud* dari kata-kata yang paling sering muncul pada data yang telah dikumpulkan pada penelitian ini:





Gambar 8. Wordcloud hasil kata yang sering muncul

Dari wordcloud di atas akan diambil tiga kata yang paling sering muncul sebagai contoh dan acuan dalam perhitungan probabilitas prior dan posterior, yakni kata "anak" yang muncul sebanyak 125 kali, "nikah" yang muncul sebanyak 37, dan "pasangan" yang muncul sebanyak 30 kali. Berikut adalah tabel dari contoh sentimen yang diambil sebagai contoh untuk dilakukan perhitungan terhadap probabilitas prior dan posterior:

Tabel 1. Hasil Klasifikasi Data

	Anak	Nikah	Pasangan	Kelas
P2	1	0	0	Pos
P62	0	0	1	Pos
P69	1	0	0	Pos
P99	0	1	0	Pos
N36	1	0	1	Neg
N48	0	1	0	Neg
N67	1	0	0	Neg
U169	1	0	0	?

3.1 Probabilitas Prior

Berikut adalah merupakan perhitungan probabilitas prior dari beberapa data yang telah dilakukan proses pengklasifikasian:

$$P(\text{Positif}) = \frac{\text{Jumlah Data Positif}}{\text{Total Jumlah Data}} = \frac{4}{7} = 0,57$$

$$P(\text{Negatif}) = \frac{\text{Jumlah Data Negatif}}{\text{Total Jumlah Data}} = \frac{3}{7} = 0,42$$

Class Positif (P(U169|Positif))

$$= P(\text{anak}=2|\text{Pos}) \times P(\text{nikah}=1|\text{Pos}) \times P(\text{pasangan}=1|\text{Pos})$$

$$= 0,5 \times 0,25 \times 0,25$$

$$= 0,03125$$

Class Negatif (P(U169|Negatif))

$$= P(\text{anak}=2|\text{Neg}) \times P(\text{nikah}=1|\text{Neg}) \times P(\text{pasangan}=1|\text{Neg})$$

$$= 0,66 \times 0,33 \times 0,33$$

$$= 0,07187$$

3.2 Probabilitas Posterior

Berikut adalah merupakan perhitungan probabilitas posterior dari beberapa data yang telah dilakukan proses pengklasifikasian:

$$\text{Posterior Positif} = P(U169|\text{Positif}) \times P(\text{Positif})$$

$$= 0,03125 \times 0,57$$

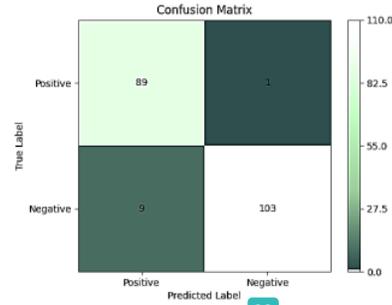
$$= 0,0178125$$

$$\text{Posterior Negatif} = P(U169|\text{Negatif}) \times P(\text{Negatif})$$

$$= 0,07187 \times 0,42$$

$$= 0,0301854$$

Setelah melakukan perhitungan nilai probabilitas tersebut, dapat diklasifikasikan bahwa data uji atau data testing ke-169 masuk ke dalam kategori sentimen negatif. Hal ini dikarenakan nilai probabilitas posterior negatif yang dihasilkan lebih besar daripada probabilitas posterior positif.



Gambar 9. Visualisasi Confusion Matrix

Berdasarkan visualisasi confusion matrix yang ditunjukkan oleh gambar 9, dilakukan perhitungan nilai *accuracy*, *recall (TPR)*, *precision (PPV)*, *TNR*, dan *NPV*. Perhitungan nilai *accuracy* dilakukan dengan tujuan untuk menilai sejauh mana algoritma pengklasifikasi Naïve Bayes dapat melakukan prediksi terhadap seluruh data uji secara benar. Selanjutnya, perhitungan nilai *recall* atau *True Positive Rate (TPR)* dilakukan untuk menilai seberapa jauh algoritma pengklasifikasi Naïve Bayes dapat melakukan deteksi terhadap seluruh sentimen positif dengan benar. Lalu, perhitungan nilai *precision* atau *Positive Predictive Value (PPV)* dilakukan untuk mengukur seberapa jauh algoritma pengklasifikasi Naïve Bayes dapat memberikan prediksi positif dengan akurat. Setelah itu, perhitungan nilai *True Negative Rate (TNR)* digunakan untuk menilai algoritma Naïve Bayes dapat seberapa jauh melakukan pendeteksian seluruh sentimen negatif secara akurat. Terakhir, yakni perhitungan nilai *Negative Predictive Value (NPV)* yang bertujuan untuk menilai seberapa jauh algoritma pengklasifikasi Naïve Bayes dapat memberikan prediksi negatif dengan akurat. Berikut adalah perhitungan dari nilai-nilai tersebut:

- **Accuracy**

$$Accuracy = \frac{89+102}{89+9+1+102}$$

$$Accuracy = 0,9502 \text{ atau } 95,02\%$$

- **Recall (True Positive Rate)**

$$Recall = \frac{89}{89+1}$$

$$Recall = 0,9889 \text{ atau } 98,89\%$$

- **Precision (Positive Predictive Value)**

$$PPV = \frac{89}{89+9}$$

$$PPV = 0,9082 \text{ atau } 90,82\%$$

- **TNR (True Negative Rate)**

$$TNR = \frac{102}{9+102}$$

$$TNR = 0,9189 \text{ atau } 91,89\%$$

- **NPV (True Negative Rate)**

$$NPV = \frac{102}{102+1}$$

$$NPV = 0,9903 \text{ atau } 99,03\%$$



Dari perhitungan tersebut dapat dijelaskan bahwa algoritma Naïve Bayes bekerja dengan baik dalam menganalisis sentimen. Hal ini dibuktikan dengan hasil perhitungan nilai *accuracy* yang mencapai angka 95,02%, yang berarti algoritma atau model klasifikasi memiliki akurasi tingkat tinggi dalam memprediksi sentimen positif dan negatif secara benar dan akurat. Selain itu, hasil perhitungan nilai *recall* (TPR) dan TNR yang hasilnya berturut-turut adalah 98,89% dan 91,89%. Hal ini berarti, nilai *recall* (TPR) dan TNR menjelaskan bahwa algoritma atau model klasifikasi mampu mendeteksi sentimen positif dan negatif secara benar. Perhitungan selanjutnya adalah perhitungan nilai *precision* (PPV) dan NPV, yang masing-masing menunjukkan hasil 90,82% dan 99,03%. Dari perhitungan tersebut dapat diketahui bahwa algoritma atau model klasifikasi mengindikasikan tiap sentimen positif dan negatif yang diprediksi adalah benar.

accuracy: 95.02%

	true Positive	true Negative	class precision
pred. Positive	89	1	98.89%
pred. Negative	9	102	91.89%
class recall	90.82%	99.03%	

Gambar 10. Hasil Pehitungan pada RapidMiner

17

4. KESIMPULAN

Dari hasil pengujian dan analisis yang telah dilakukan pada penelitian tentang analisis sentimen opini masyarakat Indonesia mengenai fenomena *childfree* yang datanya ditarik dari *Twitter* sejumlah kurang lebih 2000 data, yang kemudian data-data tersebut melalui beberapa tahap pembersihan agar lebih mudah dilakukan analisis, hingga menyisakan data sejumlah 500 data, dan akhirnya data tersebut dianalisis menggunakan algoritma pengklasifikasi Naïve Bayes dapat disimpulkan bahwa Algoritma klasifikasi Naïve Bayes yang diterapkan pada penelitian ini dinilai efektif dalam melakukan penganalisan atau pengklasifikasian sentimen opini-opini masyarakat mengenai *childfree* yang datanya dikumpulkan dari *Twitter*. Hasil nilai *accuracy* yang terhitung dari pengujian evaluasi algoritma atau model klasifikasi Naïve Bayes terhadap sentimen masyarakat mengenai *childfree* mencapai angka 95,02%, hal ini berarti bahwa algoritma pengklasifikasi Naïve Bayes dapat memprediksi sentimen-sentimen atau opini-opini masyarakat mengenai fenomena *childfree* secara benar dan akurat. Hal ini dikarenakan nilai akurasi yang dihasilkan hampir menyentuh angka sempurna. Selain itu dapat disimpulkan juga bahwa opini masyarakat mengenai fenomena *childfree* lebih condong ke arah mereka yang menanggapinya secara negatif dibandingkan positif. Hal ini dapat dibuktikan dengan sentimen positif yang berjumlah lebih banyak dari yang negatif dengan jumlah total sentimennya adalah 323 sentimen, dan sentimen negatif yang berjumlah 184 sentimen. Dengan demikian, simpulan mengenai pandangan masyarakat di Indonesia mengenai fenomena *childfree* masih dominan mengarah kepada tanggapan yang positif. Hal ini disebabkan dengan adanya pandangan mengenai pendekatan terhadap kesejahteraan pribadi dan pemahaman tentang pilihan hidup, di mana beberapa dari mereka menganggap bahwa mereka bisa mendapatkan kesejahteraan pribadi yang lebih dengan memilih untuk *childfree*. Mereka lebih menginginkan untuk fokus ke kehidupan karir, pendidikan, atau aktivitas lain yang bagi mereka dapat memberikan kebahagiaan serta kesejahteraan yang lebih dalam hidup mereka.

REFERENCES

- [1] E. Nakkerud, "Ideological Dilemmas Actualised by the Idea of Living Environmentally Childfree," *Hum. Arenas*, no. 0123456789, 2021.
- [2] D. Siregar, F. Ladayya, N. Z. Albaqi, and B. M. Wardana, "Penerapan Metode Support Vector Machines (SVM) dan Metode Naïve Bayes Classifier (NBC) dalam Analisis Sentimen Publik terhadap Konsep Child-free di Media Sosial Twitter," vol. 7, no. 1, pp. 93-104, 2023.
- [3] P. W. Ratiasadara, S. Sudarno, and T. Tarno, "Analisis Sentimen Penerapan Ppkm Pada Twitter Menggunakan Naive Bayes Classifier Dengan Seleksi Fitur Chi-Square," *J. Gaussian*, vol. 11, no. 4, pp. 580-590, 2023.
- [4] G. A. A. Jabbar Alkubaisi, S. S. Kamaruddin, and H. Husni, "Stock Market Classification Model Using Sentiment



- Analysis on Twitter Based on Hybrid Naive Bayes Classifiers,” *Comput. Inf. Sci.*, vol. 11, no. 1, p. 52, 2018.
- [5] F. N. Hasan, F. Sidik, and P. Afikah, “Sentiment Analysis of Community Response on Cooking Oil Price Increase Policy with Naïve Bayes Classifier Algorithm,” *J. Linguist. Komputasional*, vol. 5, no. 2, pp. 71–76, 2022.
- [6] M. Fahmi, A. Puspita, and Y. Yuningsih, “Sentiment Analysis of Online Gojek Transportation Services on Twitter Using the Naïve Bayes Method,” *JITK (Jurnal Ilmu Pengetah. dan Teknol. Komputer)*, vol. 8, no. 2, pp. 84–90, 2023.
- [7] H. H. Mubaroroh, H. Yasin, and A. Rusgiyono, “Analisis Sentimen Data Ulasan Aplikasi Ruangguru Pada Situs Google Play Menggunakan Algoritma Naïve Bayes Classifier Dengan Normalisasi Kata Levenshtein Distance,” *J. Gaussian*, vol. 11, no. 2, pp. 248–257, 2022.
- [8] D. Darwis, N. Siskawati, and Z. Abidin, “Penerapan Algoritma Naive Bayes Untuk Analisis Sentimen Review Data Twitter Bmkg Nasional,” *J. Tekno Kompak*, vol. 15, no. 1, p. 131, 2021.
- [9] R. Novendri, A. S. Callista, D. N. Pratama, and C. E. Puspita, “Sentiment Analysis of YouTube Movie Trailer Comments Using Naïve Bayes,” *Bull. Comput. Sci. Electr. Eng.*, vol. 1, no. 1, pp. 26–32, 2020.
- [10] R. Sari and R. Y. Hayuningtyas, “Penerapan Algoritma Naive Bayes Untuk Analisis Sentimen Pada Wisata TMI Berbasis Website,” *Indones. J. Softw. Eng.*, vol. 5, no. 2, pp. 51–60, 2019.
- [11] V. R. Prasetyo, H. Lazuardi, A. A. Mulyono, and C. Lauw, “Penerapan Aplikasi RapidMiner Untuk Prediksi Nilai Tukar Rupiah Terhadap US Dollar Dengan Metode Linear Regression,” *J. Nas. Teknol. dan Sist. Inf.*, vol. 7, no. 1, pp. 8–17, 2021.
- [12] F. Sidik, I. Suhada, A. H. Anwar, and F. N. Hasan, “Analisis Sentimen Terhadap Pembelajaran Daring Dengan Algoritma Naive Bayes Classifier,” *J. Linguist. Komputasional*, vol. 5, no. 1, p. 34, 2022.
- [13] F. Ratnawati, “Implementasi Algoritma Naive Bayes Terhadap Analisis Sentimen Opini Film Pada Twitter,” *INOVTEK Polbeng*, vol. 3, no. 1, pp. 50–59, 2018.
- [14] N. S. Wardani, A. Prahutama, and P. Kartikasari, “Analisis Sentimen Pemindahan Ibu Kota Negara Dengan Klasifikasi Naïve Bayes Untuk Model Bernoulli Dan Multinomial,” *J. Gaussian*, vol. 9, no. 3, pp. 237–246, 2020.
- [15] A. A. Farisi, Y. Sibaroni, and S. Al Faraby, “Sentiment analysis on hotel reviews using Multinomial Naïve Bayes classifier,” *J. Phys. Conf. Ser.*, vol. 1192, no. 1, 2019.
- [16] J. LING, I. P. E. N. KENCANA, and T. B. OKA, “Analisis Sentimen Menggunakan Metode Naïve Bayes Classifier Dengan Seleksi Fitur Chi Square,” *E-Jurnal Mat.*, vol. 3, no. 3, p. 92, 2014.
- [17] Pristiyono, M. Ritonga, M. A. Al Ihsan, A. Anjar, and F. H. Rambe, “Sentiment analysis of COVID-19 vaccine in Indonesia using Naïve Bayes Algorithm,” *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 1088, no. 1, p. 012045, 2021.
- [18] C. Villavicencio, J. J. Macrohon, X. A. Inbaraj, J. H. Jeng, and J. G. Hsieh, “Twitter sentiment analysis towards covid-19 vaccines in the Philippines using naïve bayes,” *Inf.*, vol. 12, no. 5, 2021.
- [19] L. Dey, S. Chakraborty, A. Biswas, B. Bose, and S. Tiwari, “Sentiment Analysis of Review Datasets Using Naïve Bayes’ and K-NN Classifier,” *Int. J. Inf. Eng. Electron. Bus.*, vol. 8, no. 4, pp. 54–62, 2016.
- [20] Z. Bokaei Nezhad and M. A. Deihimi, “Twitter sentiment analysis from Iran about COVID 19 vaccine,” *Diabetes Metab. Syndr. Clin. Res. Rev.*, vol. 16, no. 1, p. 102367, 2022.
- [21] A. Wibowo, F. N. Hasan, R. Nurhayati, and A. Wibowo, “Analisis Sentimen Opini Masyarakat Terhadap Keefektifan Pembelajaran Daring Selama Pandemi COVID-19 Menggunakan Naïve Bayes Classifier,” *J. Asimetrik J. Ilm. Rekayasa dan Inov.*, vol. 4, no. 2, pp. 239–248, 2022.
- [22] I. R. Afandi, F. H. Noor, A. A. Rizki, N. Pratiwi, and Z. Halim, “Analisis Sentimen Opini Masyarakat Terkait Pelayanan Jasa Ekspedisi Anteraja Dengan Metode Naive Bayes,” *Jlk*, vol. 5, no. 2, pp. 63–70, 2022.

ORIGINALITY REPORT

18%

SIMILARITY INDEX

16%

INTERNET SOURCES

8%

PUBLICATIONS

3%

STUDENT PAPERS

PRIMARY SOURCES

1	repository.uhamka.ac.id Internet Source	4%
2	Submitted to Universitas Brawijaya Student Paper	1%
3	journals.upi-yai.ac.id Internet Source	1%
4	www.stmik-budidarma.ac.id Internet Source	1%
5	ejurnal.seminar-id.com Internet Source	1%
6	doi.org Internet Source	1%
7	123dok.com Internet Source	1%
8	repository.itelkom-pwt.ac.id Internet Source	1%
9	sismatik.nusaputra.ac.id Internet Source	<1%

10	Tetrian Widyanto, Ina Ristiana, Arief Wibowo. "Komparasi Naïve Bayes dan SVM Analisis Sentimen RUU Kesehatan di Twitter", SINTECH (Science and Information Technology) Journal, 2023 Publication	<1 %
11	ojs.uajy.ac.id Internet Source	<1 %
12	www.djournals.com Internet Source	<1 %
13	e-journal.stmiklombok.ac.id Internet Source	<1 %
14	www.infona.pl Internet Source	<1 %
15	ejournal.itn.ac.id Internet Source	<1 %
16	ejournal3.undip.ac.id Internet Source	<1 %
17	fr.scribd.com Internet Source	<1 %
18	repository.unimus.ac.id Internet Source	<1 %
19	Muhamad Fatchan, Hari Sugeng. "Anlisa Terpilihnya Tri Rismaharini sebagai Menteri Sosial dengan Pendekatan Algorithma	<1 %

Naïve Bayes", Journal of Practical Computer Science, 2022

Publication

20

Ratih Yulia Hayuningtyas. "Penerapan Algoritma Naïve Bayes untuk Rekomendasi Pakaian Wanita", Jurnal Informatika, 2019

Publication

<1 %

21

ejurnal.teknokrat.ac.id

Internet Source

<1 %

22

nurhapsaricontoh-contoh.blogspot.com

Internet Source

<1 %

23

repository.ppns.ac.id

Internet Source

<1 %

24

Syahfira Mulya, Herry Sujaini, Tursina Tursina. "Analisis Sentimen Tren Olahraga di Masa Pandemi COVID-19 pada Twitter dengan Metode Naïve Bayes Classifier (NBC)", Jurnal Edukasi dan Penelitian Informatika (JEPIN), 2022

Publication

<1 %

25

jurnal.atmaluhur.ac.id

Internet Source

<1 %

26

repository.uin-suska.ac.id

Internet Source

<1 %

27

www.scribd.com

Internet Source

<1 %

28

Ahmad Fathan Hidayatullah, Aufa Aulia Fadila,
Kiki Purnama Juwairi, Royan Abida Nayoan.
Jurnal Linguistik Komputasional (JLK), 2019

Publication

<1 %

29

Retno Sari, Ratih Yulia Hayuningtyas.
"Penerapan Algoritma Naive Bayes Untuk
Analisis Sentimen Pada Wisata TMII Berbasis
Website", Indonesian Journal on Software
Engineering (IJSE), 2019

Publication

<1 %

30

djournals.com

Internet Source

<1 %

31

dspace.uii.ac.id

Internet Source

<1 %

32

ejurnal.stmik-budidarma.ac.id

Internet Source

<1 %

33

elibrary.nusamandiri.ac.id

Internet Source

<1 %

34

hdl.handle.net

Internet Source

<1 %

35

jtiik.ub.ac.id

Internet Source

<1 %

36

jurnal.umus.ac.id

Internet Source

<1 %

37

openlibrarypublications.telkomuniversity.ac.id

Internet Source

<1 %

38

pasardana.com

Internet Source

<1 %

39

Laurentius Dandi Andhika, Dela Regita Cahyani, Dhika Saputra, Tirta Herawati, Muhammad Khoiruddinsyah, Dedi Dwi Saputra. "Analisis Sentimen Kosumen KFC Berdasarkan Pendekatan Naive Bayes dan Ada Boost Berbasis Data Twitter", Jurnal INSAN - Journal of Information System Management Innovation, 2023

Publication

<1 %

40

Andriani Nurian. "ANALISIS SENTIMEN ULASAN PENGGUNA APLIKASI GOOGLE PLAY MENGGUNAKAN NAÏVE BAYES", Jurnal Informatika dan Teknik Elektro Terapan, 2023

Publication

<1 %

41

Anggreiny Rolangon, Axcel Weku, Green Arther Sandag. "Perbandingan Algoritma LSTM Untuk Analisis Sentimen Pengguna Twitter Terhadap Layanan Rumah Sakit Saat Pandemi Covid-19", TeIKa, 2023

Publication

<1 %

42

repository.usd.ac.id

Internet Source

<1 %

Exclude quotes On

Exclude matches Off

Exclude bibliography On