Jurnal Teknologi Informasi dan Komunikasi
Vol: 13 No 02 2022 **E-ISSN: 2477-3255**
Diterima Redaksi: 20-10-2022 | Revisi: 12-11-2022 | Diterbitkan: 30-11-2022

C4.5 Algorithm Implementation for Public Sentyment Analysis Covid-19 Vaccine

Devi Astri Nawangnugraeni[*1], M. Zakki Abdillah[2], Akrim Teguh Suseno[3]
[1,2,3] Teknologi Komputer, Fakultas Sains dan Teknologi, Institut Teknologi dan Sains Nahdlatul Ulama Pekalongan, Indonesia
[1,2,3]Jl. Karangdowo No.9 Kedungwuni Kab. Pekalongan
e-mail: [1]devinawang4@gmail.com, [2]m.zakki.abdillah@gmail.com,[3]akrim.teguh@gmail.com

Abstract
Corona virus disease is one of the dangerous diseases and has been prevented by giving vaccinations. In an effort to prevent, there must be a positive or negative public response. One of the media facilities used to convey public responses is Twitter. The public's reaction can be analyzed using public sentiment analysis using C4.5 algorithm. The purpose of paper for determine public's response to the administration of moderna and pfizer vaccinations. The implemented methodology starts from collecting data taken from tweets, pre-processing, classification using the C4.5 algorithm and validation using k-fold cross validation. Based on the results of the moderna keyword analysis, the positive sentiment response was 6% and negative sentiment was 94%, while the pfizer keyword positive sentiment was 12.4% and negative sentiment was 87.6%. The results of test iteration that have been carried out 3 times, the average error value is 38%.

Keyword: *Sentiment Public Analysis, C4.5 Algorithm, Vaccine, Covid-19*

Implementasi Algoritma C4.5 untuk Analisis Sentimen Publik Terhadap Vaksinasi Covid-19

Abstrak
Corona virus disease termasuk salah satu penyakit yang berbahaya dan sudah dilakukan pencegahan dengan memberikan vaksinasi. Dalam upaya pencegahan pasti mendapatkan respon publik baik positif maupun negatif. Fasilitas media yang digunakan untuk menyampaikan respon publik salah satunya menggunakan Twitter. Reaksi masyarakat tersebut dapat dianalisis menggunakan analisis sentimen publik dengan algoritma C4.5. Tujuan dilakukan analisis sentimen publik ini yaitu untuk mengklasifikasikan respon masyarakat terhadap pemberian vaksinasi moderna dan pfizer. Metodologi yang diimplementasikan yaitu dimulai dari pengumpulan data yang diambil dari tweet di Twitter, pre-processing, klasifikasi dengan algoritma C4.5 dan validasi dengan k-fold cross validation. Berdasarkan hasil analisis kata kunci moderna maka respon sentimen positif sebesar 6% dan untuk sentimen negatif sebesar 94%, sedangkan kata kunci pfizer sentimen positif sebesar 12,4% dan sentimen negatif

sebesar 87,6%. Hasil iterasi pengujian yang telah dilakukan sebanyak 3 kali maka rata-rata nilai eror sebesar 38%.

Kata Kunci: *Analisis Sentimen Publik, Algoritma C4.5, Vaksinasi, COVID-19*

1. Pendahuluan

Corona Virus Disease pertama teridentifikasi mulai Desember 2019 di Wuhan, China dan dikenal sebagai penyakit COVID-19. Virus ini termasuk salah satu virus yang menyebar secara cepat di Indonesia maupun di dunia. Kasus penyakit COVID-19 yang pertama di Indonesia terkonfirmasi pada bulan Maret tahun 2020 yaitu terjadi pada dua orang warga Depok [1]. Setelah kasus pertama penderita semakin banyak, menurut data Kemenkes RI bulan November 2021 jumlah penderita terkonfirmasi mencapai 4.253.598. Sampai saat ini penyebaran penyakit COVID-19 semakin meluas di Indonesia khususnya dengan adanya varian baru yaitu omicron. Hal ini sangat berdampak negatif pada semua sektor baik kesehatan, sosial, ekonomi hingga pendidikan.

Upaya pencegahan untuk menekan dampak negatif sudah dilakukan oleh pemerintah Indonesia yaitu mulai dari penerapan *lockdown* di seluruh wilayah terdampak, penggunaan masker disetiap aktivitas masyarakat, menjaga jarak dan pencegahan dengan memberikan vaksinasi kepada seluruh masyarakat Indonesia. Vaksinasi termasuk upaya pemerintah dalam mengatasi penyebaran penyakit COVID-19. Di Indonesia sebelumnya sudah menggunakan vaksin sinovac, dan sebagian besar masyarakat Indonesia sudah menggunakan vaksin sinovac. Jenis vaksin lainnya yang digunakan di Indonesia adalah pfizer dan moderna untuk mengimbangi perkembangan varian COVID-19 baru. Dalam upaya pencegahan vaksinasi ini akan mendapat reaksi publik bermacam-macam baik respon positif atau respon negatif [2]. Media sosial termasuk fasilitas yang banyak digunakan pada masyarakat dalam menyampaikan reaksi pendapat tersebut dibangingkan media lain seperti televisi, radio maupun surat kabar [3], [4]. Media yang biasa digunakan sebagai sarana penyampaian reaksi publik yaitu pada media sosial Twitter. Media sosial Twitter menjadi salah satu sumber utama karena popularitasnya sebagai media forum diskusi terkait informasi kesehatan.

Pada penelitian ini berfokus untuk melakukan analisis sentimen publik terhadap upaya pencegahan dengan vaksinasi Pfizer dan Moderna. Dengan menggunakan analisis sentimen ini bertujuan untuk mendapatkan data opini yang memiliki polaritas sehingga dapat memberikan informasi yang bersifat positif, negatif maupun netral [5], [6]. Beberapa metode-metode untuk melakukan klasifikasi yaitu *lexicon base, machine learning*, atau *hybrid approach*. Dalam hal ini metode yang digunakan untuk analisis yaitu dengan metode *machine learning*, metode tersebut dapat memprediksi polaritas sentimen publik dari data yang digunkan [5], [7].

Reaksi pendapat masyarakat ini dapat dianalisis dengan berbagai metode berdasarkan informasi yang disampaikan di Twitter tersebut. Salah satunya dengan metode analisis sentimen yaitu analisis data yang diambil dari internet yang bertujuan untuk mengetahui polaritas datanya [5]. Dengan menggunakan data polaritas maka dapat digunakan sebagai gambaran reaksi publik yaitu baik yang bersifat positif ataupun negatif [8].

Salah satu metode untuk melakukan analisis sentimen yaitu dengan mengimplementasikan metode *decision tree* dari algoritma C4.5 [9], [10], [11]. Algoritma C4.5 tersebut akan memberikan acuan untuk tingkat klasifikasi sentimen. Kelebihan algoritma ini dapat menghasilkan *decision tree* yang lebih efisien untuk data atribut bertipe diskrit atau numerik [12]. Selain itu dapat mengatasi data atribut yang hilang dan membentuk aturan yang lebih efektif untuk diimplementasikan serta kinerjanya yang lebih cepat dibandingkan dengan algoritma lainnya. Konstruksi *decision tree* pada algoritma C4.5 dilakukan dengan mengekstraksi semua data latih dan menjadikan model pengambilan keputusan berdasarkan perhitungan Gain Ratio.

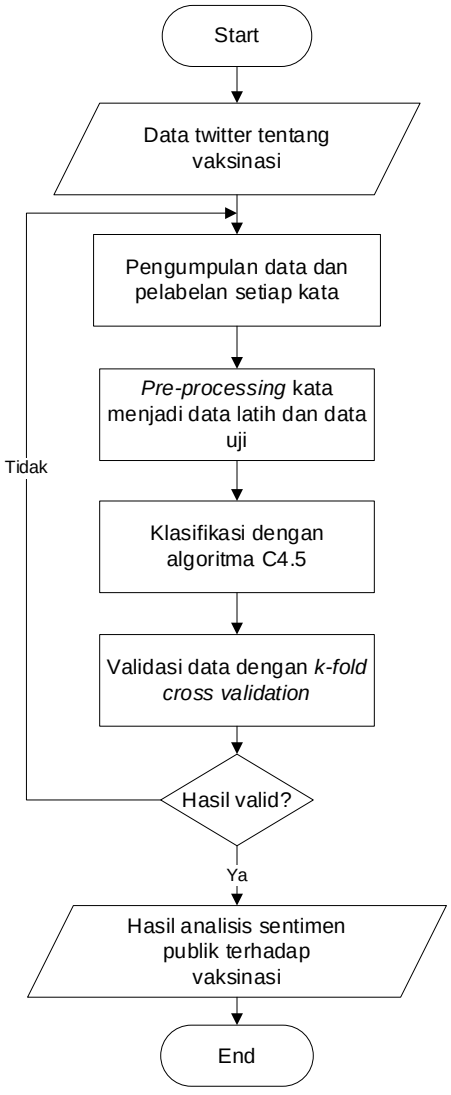
Pada penelitian sebelumnya sudah dilakukan analisis sentimen publik terhadap vaksinasi Sinovac dan Merah Putih pada media sosial Twitter menggunakan algoritma *naïve bayes* dan

https://doi.org/10.31849/digitalzone.v13i2.11658
Digital Zone is licensed under a Creative Commons Attribution International **(CC BY-SA 4.0)**

SVM [8]. Berbeda dengan penelitian sebelumnya maka penelitian ini fokus untuk mengetahui respon masyarakat terhadap vaksinasi menggunakan Pfizer dan Moderna menggunakan algoritma C4.5 melalui media sosial Twitter. Analisis ini akan menguji prediksi reaksi publik khususnya di Indonesia. Dengan menggunakan metode algoritma C4.5, maka dapat membentuk aturan *decision tree* yang lebih efektif dalam pengambilan keputusan. Hal tersebut bertujuan untuk memberikan acuan kepada pemerintah dalam menentukan jenis vaksin selanjutnya untuk mengantisipasi varian baru di dalam COVID-19.

2. Metode Penelitian

Tahapan yang dilakukan untuk analisis sentimen publik pada analisis ini yaitu pengumpulan data, pelabelan data, *preprocessing*, klasifikasi data menggunakan algoritma C4.5 kemudian dilanjutkan pengujian dengan proses validasi *k-fold cross validation*. Langkah-langkah pada analisis ini ditunjukkan pada Gambar 1.



Gambar 1. Langkah-Langkah Penelitian [8]

2.1 Pengumpulan Data dan Pelabelan Data

Berdasarkan Gambar 1 tahap ini yaitu pengumpulan data dikerjakan menggunakan teknik *web scrapping*, dengan menggunakan aplikasi Python. Data yang digunakan sebagai objek diambil mulai bulan Mei 2021 sampai bulan April 2022. Kata yang diambil menggunakan kata

kunci "pfizer", dan "moderna". Dari data yang sudah terkumpul akan diberikan pelabelan secara manual menurut sentimen negatif maupun positif.

2.2 *Pre-processing*

Tahap berikutnya yaitu data yang telah selesai dikumpulkan maka dibagikan menjadi data latih dan data uji melalui proses *pre-processing*. Hal ini berfungsi untuk mengklasifikasikan data agar lebih mudah dalam melakukan analisis. Tahapan ini dilakukan agar kalimat yang tidak terstruktur menjadi terstruktur dan menghilangkan *noise* sehingga siap untuk dilakukan analisis [5], [13].

2.3 Klasifikasi

Tahap selanjutnya yaitu setiap kata maka dilanjutkan dengan proses klasifikasi. Proses klasifikasi ini dilakukan menggunakan algoritma C4.5 yaitu sesuai dengan Persamaan (1) dan Persamaan (2) [14].

$$Entropy\ (S) = \sum_{i=1}^{n} - pi\ *\ \log_2 pi \qquad (1)$$

Dimana:
S = Kumpulan kasus
n = Banyaknya sampel
pi = Proporsi Si terhadap S

Perhitugan nilai *gain* dapat dilakukan menggunakan Persamaan (2).

$$Gain(S,A) = entropy\ (S) - \sum_{i=1}^{n} \frac{|S_i|}{|S|} * entropy\ (S_i) \quad (2)$$

Dimana:
S = Kumpulan Kasus
A = Atribut
n = Banyaknya Sampel
$|S_i|$ = Banyaknya Kasus pada partisi ke-i
$|S|$ = Banyaknya Kasus dalam S

2.4 Validasi

Tahap yang terakhir yaitu melakukan validasi data dari hasil klasifikasi menggunakan *k-fold cross validation*. Metode ini akan memisahkan data secara acak ke dalam K atau *fold*, selain itu juga dapat menghilangkan *noise* pada setiap kata agar dihasilkan akurasi yang tinggi.

3. Hasil dan Pembahasan

Hasil penelitian ini yaitu sebuah analisis sentimen publik. Analisis tersebut digunakan untuk menentukan sentimen publik terhadap vaksinasi menggunakan moderna dan pfizer dengan implementasi algoritma C4.5.

3.1 Pengumpulan Data

Data yang akan digunakan yaitu kata-kata *tweet* dari bulan Mei 2021 sampai bulan April 2022 dengan rentang wilayah Indonesia. Kata kunci yang digunakan yaitu "moderna" dan "pfizer". Data tersebut diambil menggunakan *Phyton*. Berdasarkan hasil *scraping* data *tweet* dengan kata kunci moderna sebanyak 101.880 dan data *tweet* dengan kata kunci Pfizer sejumlah 111.396. Tahap selanjutnya data hasil *scraping* akan dilakukan pelabelan yaitu sebagai penentu sentimen positif dan sentimen negatif.

3.2 *reprocessing* Data

Proses *preprocessing* ini digunakan untuk mempersiapkan data sebelum diproses ketahapan selanjutnya. Tahapan tersebut yaitu memberikan label setiap kata pada *tweet* dan diklasifikasikan ke sentiment positif dan sentiment negatif. Pada tahap *pre-processing* dilakukan dengan 5 tahapan yaitu *cleaning data*, *case folding*, normalisasi, *stemming word* dan *stopword*

https://doi.org/10.31849/digitalzone.v13i2.11658
Digital Zone is licensed under a Creative Commons Attribution International **(CC BY-SA 4.0)**

removal. Contoh hasil *preprocessing* ditunjukan pada Tabel 1.

Tabel 1. Hasil *Pre-processing* Data

Fase *Pre-processing*	**Hasil**
Kalimat dasar	abis boster moderna badan nano2 dari kemarin malam sampe siang ini minum paraetamol masih belum baikan…
Cleaning data	abis boster moderna badan nano dari kemarin malam sampe siang ini minum paraetamol masih belum baikan
Cases folding	abis boster moderna badan nano2 dari kemarin malam sampe siang ini minum paraetamol masih belum baikan
Tokenizing	‘abis’ ‘boster’ ‘moderna’ ‘badan’ ‘nano’ ‘dari’ ‘kemarin’ ‘malam’ ‘sampe’ ‘siang’ ‘ini’ ‘minum’ ‘paraetamol’ ‘masih’ ‘belum’ ‘baikan’
Normalisasi	‘habis’ ‘booster’ ‘moderna’ ‘badan’ ‘nano’ ‘dari’ ‘kemarin’ ‘malam’ ‘sampe’ ‘siang’ ‘ini’ ‘minum’ ‘parasetamol’ ‘masih’ ‘belum’ ‘baikan’
Stemming word	‘habis’ ‘booster’ ‘moderna’ ‘badan’ ‘nano’ ‘dari’ ‘kemarin’ ‘malam’ ‘sampai’ ‘siang’ ‘ini’ ‘minum’ ‘parasetamol’ ‘masih’ ‘belum’ ‘baik’
Stopword removal	‘booster’ ‘moderna’ ‘badan’ ‘kemarin’ ‘malam’ ‘siang’ ‘minum’ ‘parasetamol’ ‘belum’ ‘baik’

Pada Tabel 1 ditunjukkan tahapan fase *pro-processing* dan hasil dari setiap proses yang telah dilakukan. Proses *cleaning* data membersihkan *tweet* yang tidak berkaitan dengan vaksin moderna maupun pfizer serta menghapus duplikasi *tweet* maupun *retweet* sehingga data yang digunakan lebih akurat. Selain itu pada tahap ini akan menghapus karakter khusus dan *link* URL yang menjadi *noisy* pada kalimat *tweet*. Setelah proses *cleaning* data, data yang dihasilkan yaitu 2000 *tweet* untuk kata kunci moderna maupun pfizer. Pada fase *case folding* bertujuan untuk mengubah semua kata dengan huruf kapital pada *tweet* menjadi huruf kecil sehingga mudah diolah ditahap selanjutnya. Kemudian dilakukan normalisasi untuk merubah kata yang tidak standar menjadi kata yang biasa digunakan kehidupan sehari-hari dalam bahasa Indonesia seperti pada kata ‘abis’ diubah menjadi kata ‘habis’. Setelah itu dilakukan *stemming word* untuk merubah kata pada setiap *tweet* manjadi kata baku seperti pada kata ‘baikan’ diubah menjadi ‘baik’. Pada langkah terakhir yaitu *stopword removal* yang bertujuan menghapus kata tambahan seperti kata penghubung disetiap *tweet* seperti kata ‘habis’, ‘dari’ dan lainnya. Setelah dilakukan *pre-processing*, maka langkah selanjutnya adalah pelabelan positif dan negatif. Jumlah data *tweet* sebesar 2000 dibagi menjadi 2 data yaitu data latih dan data uji. Data yang digunakan untuk data latih sebesar 667 *tweet* untuk kata kunci moderna maupun pfizer. Sedangkan untuk

data uji yang digunakan yaitu 1333 *tweet* untuk setiap kata kunci moderna maupun pfizer. Hasil dari pelabelan data latih dapat ditunjukan pada Tabel 2.

Tabel 2. Hasil Pelabelan Data Latih

Kata Kunci	Positif	Negatif	Jumlah
Moderna	135 (20,2%)	532 (79,8%)	667
Pfizer	220 (33%)	447 (67%)	667

Pada Tabel 2 merupakan data aktual pelabelan positif maupun negatif dari kata kunci moderna dan Pfizer yang berjumlah masing-masing 667 *tweet*. Pada kata kunci moderna sentimen positif terdapat 135 *tweet* dan sentimen negatif 532 *tweet*. Sedangkan pada kata kunci pfizer sentimen positif 220 *tweet* dan 667 *tweet*. Hasil dari pelabelan data latih tersebut kata kunci moderna mempunyai dominan sentimen negatif (79,8%) dan dominan sentiment positif 20,2%. Pada kata kunci pfizer mempunyai dominan sentimen negatif 67% dan sentimen positif 33%.

3.3 Perhitungan *Information Gain*

Proses pembentukan *decision tree* dengan algoritma C4.5 terdapat lima kalimat sentimen atau opini yang diberikan kelas negatif dan kelas positif tentang vaksin moderna dan pfizer. Contoh klasifikasi kalimat opini ditunjukkan pada Tabel 3.

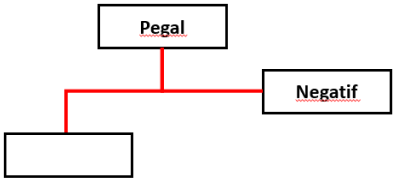
Tabel 3. Contoh Klasifikasi Kalimat Opini

No	Kalimat	Kelas
1	Adekku pfizer ga sakit, mamaku booster moderna juga ga sakit. Aman	Positif
2	Efek booster moderna lumayan kuat ya, dari kemarin pegel terus lengan kiri dan pusingg	Negatif
3	Jujur aku biasa aja yg moderna.. ya pegel normal aja alhamdulillah	Positif
4	Sakit bgt, bekas suntikkannya kaku dibadan ga enak bgt. Uda saranku jgn mau moderna	Negatif
5	2 jam setelah booster moderna masih aman	Positif

Tahapan perhitungan *gain* pada setiap atribut terhadap data *training* yang telah dilakukan berfungsi untuk menentukan atribut-atribut yang termasuk atribut penguji disetiap simpul. Sebagai contoh didapatkan lima kata sebagai atribut penguji hasil dari proses *text preprocessing* yaitu “moderna”, “pusing”, “pegal”, “sehat” dan “aman”. Tahap berikutnya yaitu pada setiap sampel tersebut dimasukan ke dalam tabel dengan kelima atribut penguji yang sudah ditentukan. Atribut penguji tersebut akan bernilai “1” apabila sampel tersebut terdapat unsur kata-kata pada atribut penguji dan akan bernilai “0” apabila pada unsur kata atribut pengujinya tidak ada. Hasil *data training* dan hasil perhitungan *information gain* pada seluruh atribut penguji ditampilkan pada Tabel 4.

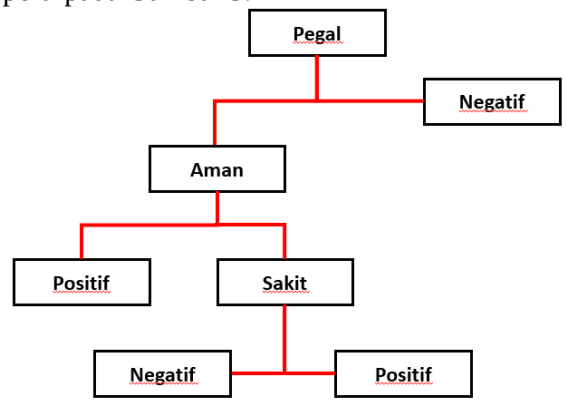
Perhitungan nilai *entropy* semua *training* data dilakukan menggunakan Persamaan 1. Setelah data *training* sudah selesai dilanjutkan dengan melakukan perhitungan nilai *gain* pada masing-masing kata dengan persamaan 2. Dari hasil perhitungan nilai gain maka dapat terbentuk simpul akar yang digunakan sebagai *rule tree*. Hasil perhitugan ditunjukkan pada Gambar 2.

https://doi.org/10.31849/digitalzone.v13i2.11658
Digital Zone is licensed under a Creative Commons Attribution International **(CC BY-SA 4.0)**



Gambar 2. Hasil Perhitungan Simpul Pegal

Gambar 2 menunjukkan hasil perhitungan simpul pada kata kunci pegal. Berdasarkan hasil tersebut, tahap berikutnya yaitu pembentukan *decision tree* maka dilanjutkan dengan membentuk data latih seperti pada Gambar 3.



Gambar 3. Hasil Perhitungan Simpul Aman dan Sakit

Gambar 3 merupakan hasil perhitungan simpul Aman dan Sakit. Hasil tersebut akan dijadikan sebagai *Decision tree.* Akhir *Decision Tree* yang dibuat menggunakan algoritma C4.5 tersebut tergantung pada *data training* yang sudah disiapkan. Kata *tweet* yang telah dilakukan klasifikasi maka akan diujikan dimulai dari simpul yang paling atas atau disebut sebagai simpul akar. Apabila salah satu *tweet* mengandung kata-kata yang termasuk atribut penguji di simpul bagian akar yaitu ‘pegal’ maka selanjutnya kata-kata atau simpul yang akan diuji yaitu simpul ruas kanan, apabila *tweet* yang tidak mengandung kata ‘pegal’ maka simpul yang selanjutnya dilakukan uji yaitu simpul pada ruas kiri. Tahapan tersebut akan dilakukan seterusnya hingga ditemukan simpul daun atau *leaf node* yang berisikan kelas mengandung sentimen negatif, sentimen positif maupun sentimen netral. Apabila terdapat *tweet* tidak mengandung kata-kata yang diujikan pada jalur tersebut, maka kata-kata *tweet* tersebut tidak dilakukan uji.

Tabel 4. Hasil Training Data

No	moderna	sakit	pegal	pusing	aman	kelas
1	1	1	0	0	1	positif
2	1	0	1	1	0	negatif
3	1	1	1	0	1	positif
4	1	1	0	0	0	negatif
5	1	0	0	0	1	positif

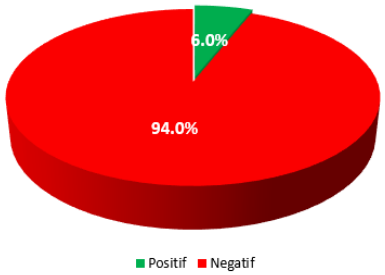
3.4 Metode Klasifikasi C4.5

Model *decision tree* yang sudah dibuat dengan algoritma C4.5 berfungsi untuk melakukan klasifikasi pada kalimat yang mengandung opini. Kalimat yang akan dilakukan pengujian yaitu kalimat yang menjadi atribut pengujian pada *decision tree*, apabila pada sebuah kalimat opini tersebut termasuk kalimat atau kata yang tidak menjadi atribut penguji pada *decision tree* maka kata tersebut tidak akan diujikan. Apabila pada kalimat tersebut terdapat kata yang menjadi atribut penguji simpul yang merupakan simpul ruas sebelah kanan dan seluruh simpul yang terdapat pada ruas kiri maka tidak akan diuji dan begitu juga sebaliknya. Data yang digunakan

untuk analisis adalah data uji sebesar 1333 pada kata kunci moderna dan Pfizer. Berikut merupakan tahapan pada klasifikasi dengan *decision tree*:

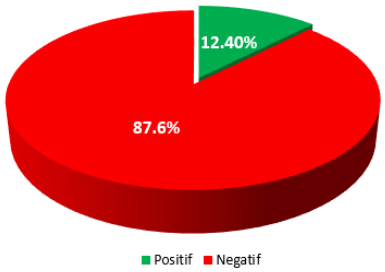
a. Simpul akar merupakan simpul yang pertama dilakukan pengujian,
b. Apabila sebuah kalimat terdapat unsur kata-kata yang termasuk atribut penguji simpul, maka simpul ruas kanan yang akan diuji terlebih dahulu. Namun jika pada sebuah kalimat tidak terdapat kata-kata yang termasuk atribut penguji simpul tersebut, maka yang akan diuji yaitu simpul ruas kiri atau ruas yang termasuk cabang dari simpul yang sebelumnya diujikan.
c. Apabila terdapat simpul yang akan diuji tetapi tidak termasuk pada *leaf node* maka yang akan dilakukan adalah langkah kedua pada simpul yang sedang diujikan.
d. Tahapan tersebut dapat selesai apabila simpul yang sedang diujikan termasuk pada *leaf node.*

Hasil prediksi dengan kata kunci moderna melalui model klasifikasi *decision tree* dengan algoritma C4.5 dapat ditunjukan pada Gambar 4. Sedangkan hasil prediksi dari kata kunci pfizer dapat ditunjukkan pada Gambar 5.



Gambar 4. Hasil Prediksi Kata Kunci Moderna

Pada Gambar 4 hasil prediksi menggunakan data uji sebesar 1333 *tweet* dan menunjukan sentimen positif masyarakat dengan kata kunci moderna sebesar 6% dan untuk sentimen negatif sebesar 94%. Hasil tersebut menunjukan bahwa tanggapan masyarakat terkait vaksin moderna cenderung lebih dominan pada sentimen negatif.



Gambar 5. Hasil Prediksi Kata Kunci Pfizer

Berdasarkan Gambar 5, data 1333 *tweet* data uji menunjukan sentimen positif dengan kata kunci pfizer sebesar 12,4% dan untuk sentimen negatif sebesar 87,6%. Hasil tersebut menunjukan bahwa tanggapan masyarakat terkait vaksin pfizer cenderung lebih dominan pada sentimen negatif.

3.5 Validasi

Proses pengujian berfungsi sebagai pengukur taraf kesesuaian hasil dari klasifikasi data sentimen yang dihasilkan dari perhitungan manual dengan hasil dari klasifikasi data dengan *decision tree* yang mengimplementasikan algoritma C4.5. Proses validasi tersebut telah dilakukan dengan *k-fold cross validation* menggunakan nilai k yaitu 3. Nilai k=3 tersebut menggambarkan jumlah *fold* datanya, dan setiap nilai subset mempunyai data dengan nilai yang berbeda. Iterasi pada pengujian yang akan dilakukan yaitu sebanyak tiga kali. Iterasi pengujian ditunjukkan pada Tabel 5.

https://doi.org/10.31849/digitalzone.v13i2.11658
Digital Zone is licensed under a Creative Commons Attribution International **(CC BY-SA 4.0)**

Tabel 5. Iterasi Pengujian

Iterasi	Data yang diuji	Data latih
1	F1	F2, F3
2	F2	F1, F2
3	F3	F1, F3

Setelah dilakukan validasi dengan iterasi sebanyak tiga kali, maka hasil iterasi pengujian menunjukkan bahwa rata-rata nilai eror pada data uji yaitu 38%. Hasil iterasi pengujian ditunjukkan pada Tabel 6.

Tabel 6. Hasil Iterasi Pengujian

Data yang diuji	Nilai eror
Iterasi ke-1	35%
Iterasi ke-2	41%
Iterasi ke-3	38%
Nilai rata-rata eror	38%

Berdasarkan hasil validasi menggunakan algoritma C4.5 maka menghasilkan rata-rata nilai eror data uji yaitu 32% dan nilai akurasi 62%. Hasil tersebut lebih tinggi yaitu 62% dibandingkan dengan penelitian sebelumnya menggunakan metode *Naive bayes* yaitu dengan hasil akurasi 61% [15]. algoritma C4.5 maka lebih efektif menghasilkan *decision tree* yang lebih efisien untuk data atribut bertipe diskrit atau numerik. Selain itu dapat mengatasi data atribut yang hilang dan membentuk aturan yang lebih efektif untuk diimplementasikan serta kinerjanya yang lebih cepat dibandingkan dengan algoritma lainnya.

4. Kesimpulan

Berdasarkan hasil penelitian yang sudah dilakukan tentang analisis sentimen publik terhadap vaksinasi Covid-19 dengan algoritma C4.5 maka dapat ditarik kesimpulan bahwa hasil analisis dominan negatif terhadap kata kunci moderna dan pfizer. Data tersebut telah dilakukan *preprocessing*, klasifikasi dan validasi. Hasil analisis kata kunci moderna respon sentimen positif sebesar 6% dan untuk sentimen negatif sebesar 94%, sedangkan kata kunci pfizer sentimen positif sebesar 12,4% dan sentimen negatif sebesar 87,6%. Hasil analisis pengujian yang sudah dilakukan menghasilkan rata-rata nilai eror adalah 38%.

Daftar Pustaka

[1] A. Susilo et al., "Coronavirus Disease 2019: Tinjauan Literatur Terkini," J. Penyakit Dalam Indones., vol. 7, no. 1, pp. 45–67, 2020, doi: 10.7454/jpdi.v7i1.415.

[2] S. Yousefinaghani, R. Dara, S. Mubareka, A. Papadopoulos, and S. Sharif, "An analysis of COVID-19 vaccine sentiments and opinions on Twitter," Int. J. Infect. Dis., vol. 108, pp. 256–262, 2021, doi: 10.1016/j.ijid.2021.05.059.

[3] L. Y. C. Wong and J. Burkell, "Motivations for sharing news on social media," ACM Int. Conf. Proceeding Ser., vol. Part F1296, 2017, doi: 10.1145/3097286.3097343.

[4] Y. Shi et al., "Knowledge and attitudes of medical staff in Chinese psychiatric hospitals regarding COVID-19," Brain, Behav. Immun. - Heal., vol. 4, no. March, p. 100064, 2020, doi: 10.1016/j.bbih.2020.100064.

[5] V. K. S. Que, A. Iriani, and H. D. Purnomo, "Analisis Sentimen Transportasi Online Menggunakan Support Vector Machine Berbasis Particle Swarm Optimization," J. Nas. Tek. Elektro dan Teknol. Inf., vol. 9, no. 2, pp. 162–170, 2020, doi: 10.22146/jnteti.v9i2.102.

[6] M. A. Fauzi, "Random forest approach fo sentiment analysis in Indonesian language," Indones. J. Electr. Eng. Comput. Sci., vol. 12, no. 1, pp. 46–50, 2018, doi: 10.11591/ijeecs.v12.i1.pp46-50.
[7] H. Tuhuteru and A. Iriani, "Analisis Sentimen Perusahaan Listrik Negara Cabang Ambon Menggunakan Metode Support Vector Machine dan Naive Bayes Classifier," J. Inform. J. Pengemb. IT, vol. 3, no. 3, pp. 394–401, 2018, doi: 10.30591/jpit.v3i3.977.
[8] B. Laurensz and Eko Sediyono, "Analisis Sentimen Masyarakat terhadap Tindakan Vaksinasi dalam Upaya Mengatasi Pandemi Covid-19," J. Nas. Tek. Elektro dan Teknol. Inf., vol. 10, no. 2, pp. 118–123, 2021, doi: 10.22146/jnteti.v10i2.1421.
[9] M. I. Dualingga, E. Budianita, M. Fikry, M. Affandes, and F. Yanto, "Klasifikasi Komentar Terhadap Vaksin Covid-19 Menggunakan Algoritma C4 . 5 Pada Media Sosial Instagram," no. November, pp. 24–31, 2021.
[10] M. R. Fajriansyah and Siswanto, "Analisis Sentimen Pengguna Twitter Terhadap Partai Politik Pendukung Calon Gubernur Di Jakarta Menggunakan Algoritma C4 . 5 Decision Tree Learning," Skanika, vol. 1, no. 2, pp. 697–703, 2018, [Online]. Available: https://jom.fti.budiluhur.ac.id/index.php/SKANIKA/article/view/278
[11] A. Sutedi, H. Aulawi, E. Walujodjati, D. Destiani, and S. Fatimah, "C4 . 5 Algorithm for Disaster Identifier System," J. Tek. Inform., vol. 3, no. 3, pp. 1–6, 2022.
[12] B. Hermanto, A. S. Azhari SN, and F. P. Putra, "Analisis Kinerja Decision Tree C4.5 dalam Prediksi Potensi Pelunasan Kredit Calon Debitur," INOVTEK Polbeng - Seri Inform., vol. 2, no. 2, p. 189, 2017, doi: 10.35314/isi.v2i2.206.
[13] H. Irsyad, A. Farisi, and M. R. Pribadi, "Klasifikasi Opini Masyarakat Terhadap Jasa ISP MyRepublic dengan Naïve Bayes," J. Nas. Tek. Elektro dan Teknol. Inf., vol. 8, no. 1, p. 30, 2019, doi: 10.22146/jnteti.v8i1.487.
[14] Y. I. Kurniawan, "Perbandingan Algoritma Naive Bayes dan C.45 dalam Klasifikasi Data Mining," J. Teknol. Inf. dan Ilmu Komput., vol. 5, no. 4, p. 455, 2018, doi: 10.25126/jtiik.201854803.
[15] Y. P.T, "Analisis Sentimen Vaksin Covid19 Menggunakan Naive Bayes," 2022.

Digital Zone is licensed under a Creative Commons Attribution International **(CC BY-SA 4.0)**