

Klasifikasi Komentar Terhadap Vaksin Covid-19 Menggunakan Algoritma C4.5 Pada Media Sosial Instagram

M. Ikhsan Dualingga¹, Elvia Budianita², M. Fikry³, Muhammad Affandes⁴, Febi Yanto⁵

Teknik Informatika UIN Sultan Syarif Kasim Riau
Jl. Subrantas Km. 15, Pekanbaru, 28293, 0761-562223

e-mail: m.ikhsan.dualingga@students.uin-suska.ac.id, elvia.budianita@uin-suska.ac.id

Abstrak

Berdasarkan survei mengenai penerimaan vaksin COVID-19 yang dilakukan oleh Kementerian Kesehatan (Kemenkes) Republik Indonesia, Indonesian Technical Advisory Group on Immunization (ITAGI), United Nations Children's Fund (UNICEF), dan World Health Organization (WHO) yang dilakukan pada September 2020 dan melibatkan 115.000 responden, memperoleh hasil bahwa masih banyak masyarakat yang ragu bahkan menolak vaksinasi COVID-19, di mana sebanyak 7,6% menolak dan 27% ragu-ragu. Penyebab penolakan dan keraguan mengenai vaksin tersebut beragam, diantaranya tidak yakin terhadap keamanan vaksin, ragu terhadap efektivitas vaksin, takut terhadap efek samping vaksin, tidak mempercayai kegunaan vaksin, dan karena keyakinan agama. Tanggapan masyarakat banyak dituangkan pada media sosial seperti Instagram sehingga penelitian ini dilakukan untuk menganalisa komentar positif dan tidak positif masyarakat terhadap vaksin menggunakan metode *text mining* dan klasifikasi Algoritma C4.5. Berdasarkan hasil pengujian dengan perbandingan 70:30, 80:20, 90:10 pada 1200 komentar diperoleh hasil akurasi terbaik 90% dengan nilai recall terbaik adalah 96,47%.

Kata Kunci: Algoritma C4.5, *Confusion Matrix*, Covid-19, Instagram, Vaksin

Abstract

Based on a survey on receipt of COVID-19 vaccines conducted by the Kementerian Kesehatan (Kemenkes) of the Republic of Indonesia, the Indonesian Technical Advisory Group on Immunization (ITAGI), United Nations Children's Fund (UNICEF), and the World Health Organization (WHO) conducted in September 2020 and involving 115,000 respondents, the results show that there are still many people who are hesitant and even refuse the COVID-19 vaccination, of which 7.6% refuse and 27% are hesitant. The reasons for refusal and doubt about the vaccine are various, including not believing in the safety of the vaccine, doubting the effectiveness of the vaccine, fearing the side effects of the vaccine, not believing in the usefulness of the vaccine, and because of religious beliefs. Many people's responses are poured on social media such as Instagram, so this research was conducted to analyze the positive and non-positive comments of the public on vaccines using text mining methods and the classification of the C4.5 Algorithm. Based on the test results with a comparison of 70:30, 80:20, 90:10 on 1200 comments, the best accuracy result is 90% with the best recall value is 96.47%.

Keywords: C4.5 Algorithm, Covid-19, Instagram, Text mining, Vaccines

1. Pendahuluan

Pada awal pengembangan vaksin Covid-19 pertama, masyarakat masih enggan mengikuti vaksinasi karena khawatir akan keamanan dan efektivitas kerja vaksin tersebut. Banyak dari masyarakat yang tidak mempercayai penggunaan vaksin sebagai solusi dalam mengakhiri pandemi. Berdasarkan survei mengenai penerimaan vaksin COVID-19 yang dilakukan oleh Kementerian Kesehatan (Kemenkes) Republik Indonesia, Indonesian Technical Advisory Group on Immunization (ITAGI), United Nations Children's Fund (UNICEF), dan World Health Organization (WHO) yang dilakukan pada September 2020 dan melibatkan 115.000 responden, memperoleh hasil bahwa masih banyak masyarakat yang ragu bahkan menolak vaksinasi COVID-19, di mana sebanyak 7,6% menolak dan 27% ragu-ragu. Penyebab penolakan dan keraguan mengenai vaksin tersebut beragam, diantaranya tidak yakin terhadap keamanan vaksin, ragu terhadap efektivitas vaksin, takut terhadap efek samping vaksin, tidak mempercayai kegunaan vaksin, dan karena keyakinan agama (SatgasCOVID-19, 2020b) [1].

Masyarakat banyak memberikan tanggapan salah satunya di media sosial terhadap vaksinasi yang akan dilakukan. Sebanyak 170 juta pengguna media sosial di Indonesia mengaksesnya pada Januari 2021, ini mengalami peningkatan 6,3 persen dari 2020 sampai 2021. Berbagai macam media sosial yang dikunjungi oleh pengguna di Indonesia salah satunya Instagram dengan mencakup lebih dari 86 persen dari total pengguna [2]. Media sosial Instagram berkembang sebagai media informasi edukasi. Pada penelitian Sari dan Basit, 2020 memperoleh kesimpulan bahwa penggunaan media sosial Instagram sebagai media edukasi khususnya pembelajaran tentang *parenting* untuk mencari informasi mengenai *parenting* oleh followers akun Instagram @parentalk.id ini mampu memberikan sebuah respons kognitif (informasi), afektif (emosi) dan tindakan behavioral (tindakan) dalam mengakses, melihat, dan mencari postingan akun Instagram [3].

Terdapat metode analisis yang dapat digunakan untuk menganalisis pendapat masyarakat berdasarkan informasi yang ada pada media sosial seperti Instagram. Salah satu diantaranya adalah metode analisis sentimen. Metode analisis sentimen merupakan salah satu metode untuk menganalisis data yang diperoleh dari internet sehingga dapat diketahui polaritas dari data tersebut. Polaritas dari opini yang ada dapat dikumpulkan, sehingga akan dapat digunakan untuk memprediksi suasana public atau gambaran perasaan netizen bersifat negatif atau positif [4].

Penelitian mengenai analisis sentiment terhadap media sosial telah dilakukan dengan berbagai metode. Seperti pada penelitian yang dilakukan oleh Muhammad Iqbal Aditama, Rizqeya Irfan Pratama, Kevin Hafizzana Untoro Wiwaha, Nur Aini Rakhmawati dengan judul "Analisis Klasifikasi Sentimen Pengguna Media Sosial Twitter Terhadap Pengadaan Vaksin COVID-19 (2020)" memberikan hasil dengan menggunakan klasifikasi polaritas text persentase opini masyarakat terhadap vaksin corona yaitu 48% positif, 29% netral, dan 23% negative [5].

Sedangkan penelitian yang dilakukan oleh Brian Laurenz dan Eko Sedyono, 2021 melakukan pengujian dengan jumlah data yang digunakan sebanyak 845 tweet, dengan menggunakan dua kata kunci, yaitu "vaksinmerahputih" dan "vaksinsinovac". Data kemudian dibagi menjadi 253 data latih dan 592 data uji. Klasifikasi dilakukan menggunakan metode SVM dan Naïve Bayes. Hasil klasifikasi dari metode Naïve Bayes mendapatkan rata-rata akurasi 85,59%, sedangkan SMV sebesar 84,41%. Hasil sentimen pada metode Naïve Bayes dengan kata kunci "vaksinsinovac" mendapatkan sentimen positif 66% dan negatif 34%, sedangkan "vaksinmerahputih" memperoleh sentimen positif 89% dan negatif 11%. Metode SVM dengan kata kunci "vaksinsinovac" mendapatkan sentimen positif 96% dan negatif 4%, sedangkan "vaksinmerahputih" mendapatkan sentimen positif 98% dan negative 2% [6].

Penelitian selanjutnya dilakukan oleh Zeli Erika Sholikha, dkk yakni analisa sentimen pengguna e-money pada twitter menggunakan algoritma C4.5 dan naïve bayes. Uji coba dilakukan dengan menggunakan data testing sebanyak 90 kalimat. Setiap data diklasifikasikan dengan label positif, negatif dan undefined. Performa dari hasil metode Naïve Bayes dan C4.5 mencapai akurasi 80% dan 84% dengan akurasi *recall* pada metode C4.5 mencapai 100% [7]. Algoritma ini sudah banyak digunakan dalam implementasi penentuan keputusan karena memiliki banyak kelebihan. Salah satu kelebihan ini misalnya dapat mengolah data numerik dan diskret, dapat menangani nilai atribut yang hilang, menghasilkan aturan-aturan yang mudah diinterpretasikan dan performanya merupakan salah satu yang tercepat dibandingkan dengan algoritma lain.

Berdasarkan permasalahan tersebut, maka penelitian ini akan menerapkan algoritma C4.5 dalam mengklasifikasikan sentiment positif dan tidak positif masyarakat terhadap vaksin covid-19 berdasarkan komentar di media sosial Instagram serta mengetahui akurasi klasifikasi yang dihasilkan metode C4.5.

2. Metode Penelitian

Metode penelitian terdiri atas beberapa tahapan diantaranya :

1. Rumusan masalah yaitu bagaimana menerapkan algoritma C4.5 untuk melakukan klasifikasi komentar positif dan tidak positif terhadap komentar mengenai vaksin covid-19 pada media sosial *Instagram*.

2. Pengumpulan Data. Data yang digunakan pada penelitian ini merupakan data komentar postingan dari akun Instagram Kemenkes RI yang berhubungan dengan vaksin covid-19. Data dikumpulkan dengan bantuan dari situs *Export Comments*, kemudian data csv diambil sebanyak 1200 komentar pada postingan yang berkaitan dengan vaksin covid-19. Dataset akan diambil secara imbang berdasarkan pembagian kelas, 600 untuk kelas positif dan 600 untuk kelas tidak positif. Setelah itu dilakukan pelabelan secara manual terhadap data latih oleh seorang psikolog untuk menentukan kelas positif dan tidak positif. Komentar positif merupakan suatu pendapat yang memiliki arti baik dan mampu mengajak pembaca untuk berfikir positif terhadap suatu persoalan seperti pujian dan dukungan. Sedangkan komentar tidak positif merupakan suatu pendapat yang tidak memiliki arti baik atau tidak berpihak ke suatu persoalan.

Contoh kalimat positif : “ aman dan mantaap suntik vaksin kok kawan, sertifikat juga dapat dan terima kasih.”

Contoh kalimat tidak positif : “gimana sih ni zaman, jelas" cina udah selesai d negaranya, nah d negara ini masi trending ni virus. Pake d vaksin" segala. Sementara negara sana aja kakak pake vaksinlah, jelas sekali ni vaksin sungguh meragukan”.

3. Tahap analisa yang terdiri atas analisa kebutuhan data, text pre-processing dan analisa klasifikasi menggunakan algoritma C4.5.
4. Perancangan ini terdiri dari perancangan *database* dan *interface*.
5. Implentasi dan Pengujian. Pada penelitian ini, pengujian dilakukan terhadap data komentar tidak menggunakan emoji dan yang menggunakan emoji dengan perbandingan data latih dan data uji 70:30, 80:20, 90:10. Kemudian dihitung akurasi tersebut menggunakan *confusion matrix*. Pengujian emoji dilakukan untuk mengetahui seberapa besar pengaruh emoji pada analisis sentiment.

2.1. Analisa Kebutuhan Data

Pengumpulan data menggunakan tools *Export Comments* yaitu dengan menyalin link pada sebuah postingan di akun yang telah ditentukan lalu dipindahkan pada tools Export Comment dan klik tombol download file lalu secara otomatis data komentar yang ada pada postingan tersebut akan diambil dalam format csv.

2.2. Tahapan Text Pre-Processing

Pra-proses (Pre-processing) data merupakan salah satu langkah penting dalam melakukan suatu analisa klasifikasi yang berfungsi untuk membuang data dari unsur – unsur yang tidak dibutuhkan untuk mempercepat proses klasifikasi. Data komentar yang Panjang akan tetap dimasukkan sebagai dataset, tidak dipisah sebagai 2 kalimat. Berikut ini adalah tahapan pra-proses data yang digunakan:

1. *Case Folding*. Tahapan case folding merupakan proses untuk mengubah seluruh huruf menjadi huruf kecil. Berikut tabel dibawah ini adalah contoh hasil data yang telah dilakukan proses case folding.
2. *Tokenizing*. Tahapan ini memisah kalimat menjadi perkata dan termasuk karakter abjad atau numberik yang telah dipisahkan oleh spasi dan tanda baca.
3. *Cleaning*. Tahapan ini menghilangkan tanda baca, emoji, angka, dan link url.
4. *Normalization*. Tahapan ini merupakan perubahan dari kata tidak baku menjadi kata baku yang sesuai dengan ejaan kamus besar Bahasa Indonesia. Hasil dari proses ini diperoleh dengan kamus kosakata IndonesianKataBakuChecker.
5. *Stopword Removal*. Tahapan ini kata yang tidak berguna atau kata tanpa ketergantungan dengan suatu topik akan dihapus. Proses ini akan dilakukan dengan menggunakan library sastrawi. Kata yang dihapus berupa kata yang tidak berguna atau kata tanpa ketergantungan dengan suatu topik seperti kata saya, mau, di, sama, harus, yang, tidak, ada, karena, itu, dari, dululah, ini, karna, juga, untuk, setelah, dan, agar, segera, tetap, saja, ikuti, tersebut, dan cukup

6. *Stemming*. Proses stemming merupakan proses menghapus kata-kata imbuhan, awalan, tengah, dan, akhiran pada kata dasar dengan menggunakan library Sastrawi.
7. *Text Transformation* atau pembobotan adalah proses perubahan token teks menjadi nilai sebuah bobot yang menjadi numerik. Proses ini menggunakan pembobotan TF-IDF. Tahapan-tahapan yang dilakukan pada proses ini yaitu menghitung *document frequency* (df), *term frequency* (tf), dan *inverse document frequency* (idf). Pada tabel 1 menunjukkan contoh hasil pembobotan tf-idf pada 6 dokumen.

Tabel 1. Contoh Hasil Pembobotan Kata Setelah Tahapan *Stemming*

Term/Kata	D1	D2	D3	D4	D5	D6	df	idf Log(n/df)
bapak	1		1				2	0,477
vaksin	2	1	2	1	2		8	-0,125
malam	1						1	0,778
kumpul	2						2	0,477
teman	4					1	5	0,079
prokes	1						1	0,778
kangen	1						1	0,778

8. *Feature Selection*. Tahapan ini adalah memilih fitur – fitur dari semua token yang telah ditransformasi dan mempunyai nilai bobot idf seperti pada tahap sebelumnya. Tahap ini dilakukan karena untuk mengurangi jumlah fitur dan memilih token – token tertentu sehingga token tersebut dapat mewakili data dari dokumen tersebut. Pada penelitian ini yang akan diterapkan adalah seleksi bobot $0 < idf \leq 1$, namun jika kondisi tersebut tidak terpenuhi maka cukup ambil 10% dari nilai $idf > 1$. Pada penelitian ini proses features selection yang dilakukan dengan seleksi bobot adalah $df \geq 2$, karena jumlah data sedikit. Sebagai contoh pada tabel 1, maka kata “bapak (A1), vaksin (A2), kumpul (A3), teman (A4)” dijadikan parameter selanjutnya untuk proses klasifikasi pada metode C4.5.

2.3. Analisa Penyelesaian menggunakan Algoritma C4.5

Setelah tahap *text pre-processing*, maka selanjutnya melakukan klasifikasi text menggunakan algoritma C4.5.

1) Algoritma C4.5

Berdasarkan analisa kebutuhan data pada tahap yang telah dilakukan sebelumnya, maka pada tahap ini akan menjelaskan cara yang akan digunakan pada klasifikasi data. Dalam menentukan pohon keputusan diambil pada pemilihan atribut yang memiliki nilai prioritas tertinggi atau mempunyai nilai *gain* tertinggi berdasarkan nilai *entropy* atribut tersebut sebagai dasar atribut klasifikasi. Kemudian secara bertahap cabang-cabang pohon dibentuk sehingga seluruh pohon terbentuk. *Entropy* adalah jumlah data yang tidak relevan dengan informasi dari kumpulan data. *Gain* adalah informasi yang didapat dari perubahan *entropy* pada suatu kumpulan data, baik melalui observasi atau dengan melakukan partisipasi terhadap suatu set data.

Persamaan dari rumus mencari entropy yaitu [8].

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

Dimana:

- S : himpunan kasus
- n : jumlah partisi S
- pi : proporsi dari Si terhadap S

Sedangkan mencari nilai Gain (S,A) dengan persamaan :

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

Dimana:

- S : himpunan kasus
- A : atribut
- n : jumlah partisi atribut A

$|S_i|$: jumlah kasus pada partisi ke- i
 $|S|$: jumlah kasus dalam S

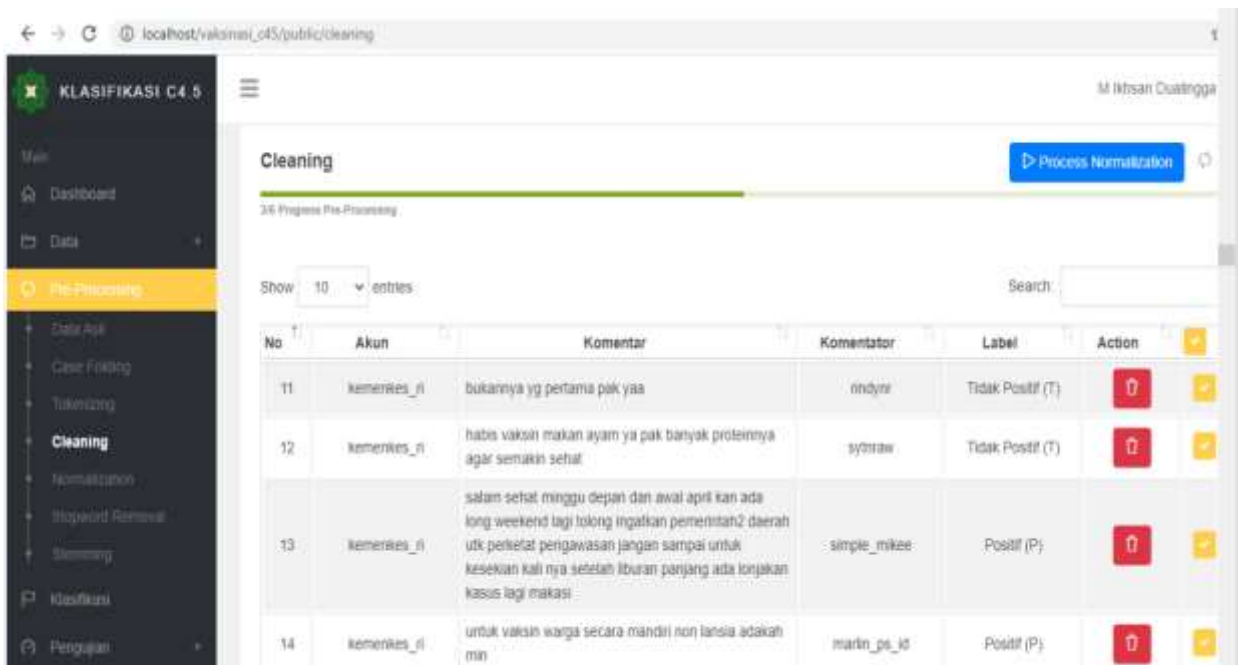
Tabel 2. Hasil perhitungan *Entropy* dan *Gain*

Node		Jumlah (S)	Positif	Tidak Positif	Entropy	Gain
1	Total	6	3	3	1	0
	A1					
		Ya	2	1	1	0,1909
		Tidak	4	2	1	
	A2					0,4591
		Ya	5	2	3	
		Tidak	1	1	1	0,1909
	A3					
	Ya	2	2	0	0,8113	
	Tidak	4	1	3		
A4					0,9719	
	Ya	1	1	0		
	Tidak	5	2	3	0,9719	

3. Hasil dan Pembahasan

Setelah melakukan analisa kebutuhan data, text pre-processing dan metode, maka selanjutnya adalah melakukan pengujian terhadap metode C4.5. Pengujian algoritma C4.5 diterapkan dengan pengujian keakuratan prediksi sistem menggunakan *confusion matrix*. Pengujian terbagi menjadi 2 pembagian data, yaitu data komentar tidak menggunakan emoji dan data komentar dengan mempertahankan emoji. Penggunaan *confusion matrix* ialah untuk menghitung akurasi sebuah metode dalam proses klasifikasi data dengan membandingkan antara hasil dari yang di prediksi oleh sistem terhadap nilai yang sebenarnya.

Gambar 2 menunjukkan implementasi klasifikasi sentiment terhadap vaksin menggunakan algoritma C4.5 dan tahapan *text pre processing*.



Gambar 1. Implementasi Klasifikasi *Sentiment* Terhadap Vaksin menggunakan Algoritma C4.5 dan Tahapan *Text Pre Processing*

3.1 Pengujian Data Komentar Tidak Menggunakan Emoji

Proses pengujian komentar menggunakan emoji, dilakukan dengan membagi data yang terdiri atas data latih dan data uji menjadi 70:30, 80:20, 90:10.

1. Pengujian Perbandingan 70:30

Pada pengujian dengan perbandingan 70:30 artinya data latih berjumlah 840 dan data uji berjumlah 360.

Tabel 3. Pengujian Perbandingan 70:30

Aktual \ Prediksi	Prediksi		Total data	Akurasi	Precision	Recall
	positif	Tidak positif				
positif	130	6	136			
Tidak positif	31	193	224	89.72%	80.75%	95.59%
Total Prediksi	161	199				

Data komentar yang berlabel positif dikenali dengan benar berjumlah 130 sedangkan data komentar yang berlabel tidak positif berjumlah 193.

Menghitung akurasi, *precision*, dan *recall* dapat dilihat pada perhitungan di bawah ini:

a. Akurasi

Menghitung akurasi menggunakan persamaan

$$\frac{(130 + 193)}{(360)} * 100\% = 89.72\%$$

b. Precision

Menghitung akurasi menggunakan persamaan

$$\frac{130}{31 + 130} * 100\% = 80.75\%$$

c. Recall

Menghitung akurasi menggunakan persamaan

$$\frac{130}{6 + 130} * 100\% = 95.59\%$$

2. Pengujian akurasi pada perbandingan data 80:20

Pada pengujian dengan perbandingan 80:20 artinya data latih berjumlah 960 dan data uji berjumlah 240.

Tabel 4. Pengujian Perbandingan 80:20

Aktual \ Prediksi	Prediksi		Total Data	Akurasi	Precision	Recall
	Positif	Tidak Positif				
Positif	82	3	85			
Tidak Positif	21	134	155	90.00%	76.61%	96.47%
Total Prediksi	103	137				

Data komentar yang berlabel positif dikenali dengan benar berjumlah 82 sedangkan data komentar yang berlabel tidak positif berjumlah 134.

3. Pengujian akurasi pada perbandingan data 90:10

Pada pengujian dengan perbandingan 80:20 artinya data latih berjumlah 1080 dan data uji berjumlah 120.

Tabel 5. Pengujian Perbandingan 90:10

Aktual \ Prediksi	Prediksi		Total Data	Akurasi	Precision	Recall
	Positif	Tidak Positif				
Positif	32	4	36			
Tidak Positif	10	74	84	88.33%	76.19%	88.89%
Total Prediksi	42	78				

Data komentar yang berlabel positif dikenali dengan benar berjumlah 32 sedangkan data komentar yang berlabel tidak positif berjumlah 74.

Pada setiap pengujian perbandingan data pada komentar yang tidak menggunakan emoji terlihat bahwa setiap data aktual tidak positif lebih banyak dikenali sebagai komentar positif.

3.2 Pengujian Data Komentar menggunakan Emoji

Selanjutnya pengujian dengan mempertahankan emoji pada data komentar. Pengujian menggunakan emoji ini tidak dihapusnya emoji yang ada pada data komentar, lalu emoji diubah menyesuaikan arti dari emoji tersebut. Seperti emoji 😊 artinya senyum. Berikut adalah hasil dari pengujianya. Jumlah data dan perbandingan tetap sama dengan yang tidak menggunakan emoji.

1. Hasil dari perbandingan data 70:30

Pengujian pertama menggunakan perbandingan 70:30 dari 1200 data, terbentuk data uji berjumlah 360 dan data latih 840 data.

Tabel 6. Pengujian Perbandingan 70:30

Aktual \ Prediksi	Prediksi		Total data	Akurasi	Precision	Recall
	positif	Tidak positif				
positif	129	7	136	90.00%	81.65%	94.85%
Tidak positif	29	195	224			
Total Prediksi	158	202				

Pada pengujian 70:30 ini data komentar yang berlabel positif dikenali dengan benar berjumlah 129 sedangkan data komentar yang berlabel tidak positif berjumlah 195.

2. Hasil dari perbandingan data 80:20

Pengujian kedua menggunakan perbandingan 80:20 dari 1200 data, terbentuk data uji berjumlah 240 dan data latih 960 data.

Tabel 7. Pengujian Perbandingan 80:20

Aktual \ Prediksi	Prediksi		Total data	Akurasi	Precision	Recall
	positif	Tidak positif				
positif	81	4	85	90.00%	80.20%	95.29%
Tidak positif	20	135	155			
Total Prediksi	101	139				

Pada pengujian 80:20 ini data komentar yang berlabel positif dikenali dengan benar berjumlah 81 sedangkan data komentar yang berlabel tidak positif berjumlah 135.

3. Hasil dari perbandingan data 90:10

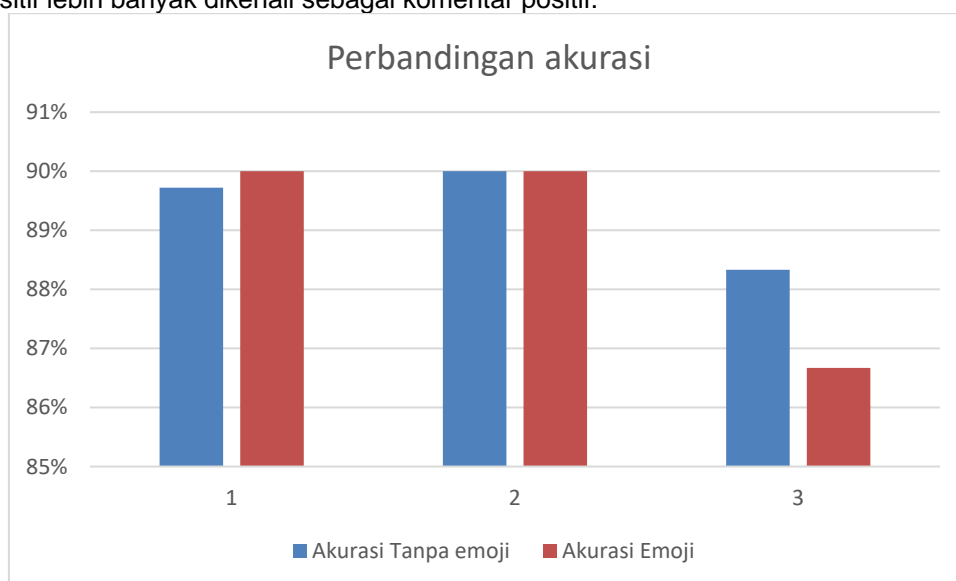
Pengujian selanjutnya menggunakan perbandingan 90:10 dari 1200 data, terbentuk data uji berjumlah 120 dan data latih 1080 data.

Tabel 8. Pengujian Perbandingan 90:10

Aktual \ Prediksi	Prediksi		Total data	Akurasi	Precision	Recall
	positif	Tidak positif				
positif	29	7	36	86.67%	76.32%	80.56%
Tidak positif	9	75	84			
Total Prediksi	38	82				

Pada pengujian 90:10 ini data komentar yang berlabel positif dikenali dengan benar berjumlah 29 sedangkan data komentar yang berlabel tidak positif berjumlah 75. Berdasarkan hasil pengujian diperoleh bahwa akurasi menggunakan emoji dan tanpa emoji memberikan akurasi yang sama yaitu 90% pada pembagian data 80:20. Pada gambar 2 dapat dilihat perbandingan akurasi berdasarkan pembagian data 1 (70:30), 2 (80:20), dan 3 (90:10). Pada setiap pengujian perbandingan data pada komentar baik yang tidak

menggunakan emoji maupun menggunakan emoji terlihat bahwa setiap data aktual tidak positif lebih banyak dikenali sebagai komentar positif.



Gambar 2. Perbandingan Akurasi pada Pengujian menggunakan Emoji dan Tanpa Emoji

4. Kesimpulan

Berdasarkan dari hasil pengujian yang di lakukan pada penelitian ini, dapat diambil beberapa kesimpulan sebagai berikut :

1. Algoritma C4.5 dapat diterapkan untuk klasifikasi sentimen masyarakat terhadap vaksin covid-19 berdasarkan komentar di Instagram.
2. Berdasarkan hasil pengujian diperoleh bahwa akurasi menggunakan emoji dan tanpa emoji memberikan akurasi yang sama yaitu 90% pada pembagian data 80:20.
3. Berdasarkan klasifikasi yang telah dilakukan dengan jumlah data komentar 1200 menggunakan algoritma C4.5 dari tiga pembagian data uji dan data latih diperoleh kesimpulan komentar tidak positif lebih banyak dikenali sebagai komentar positif.

Daftar Pustaka

- [1] Nugroho.,A., S. dan Hidayat., N., I., "Efektivitas dan Kemanan Vaksin Covid-19 : Studi Referensi", Jurnal Keperawatan Profesional (JKP), Volume 9, Nomor 2 Agustus 2021.
- [2] wearesosial, "DIGITAL 2021: INDONESIA," <https://wearesosial.com/digital-2021>, 2021. [Online]. Available: <https://datareportal.com/reports/digital-2021-indonesia>.
- [3] F. F. Rachman and S. Pramana, "Analisis Sentimen Pro dan Kontra Masyarakat Indonesia tentang Vaksin COVID-19 pada Media Sosial Twitter," *Heal. Inf. Manag. J. ISSN*, vol. 8, no. 2, pp. 2655–9129, 2020.
- [4] D. H. Kamagi and S. Hansun, "Implementasi Data Mining dengan Algoritma C4.5 untuk Memprediksi Tingkat Kelulusan Mahasiswa," *J. Ultim.*, vol. 6, no. 1, pp. 15–20, 2014, doi: 10.31937/ti.v6i1.327.
- [5] Aditama.,I.,M., Pratama., I., R., Wiwaha.,U., H., K., Rakhmawati., A., "Analisis Klasifikasi Sentimen Pengguna Media Sosial Twitter Terhadap Pengadaan Vaksin COVID-19", *JIEET*, Vol 04, No. 02, 2020 , e-ISSN 2549 -869X
- [6] Laurenz.,B.,Pratama., Sedyono., E., "Analisis Sentimen Masyarakat terhadap Tindakan Vaksinasi dalam Upaya Mengatasi Pandemi Covid-19", *Jurnal Nasional Teknik Elektro dan Teknologi Informasi*, Vol 10, No. 02, 2021 , e-ISSN 2301-4156
- [7] Sholikha.,Z., E., Puspaningrum., E., Y., dan JS., S., W., " Analisa sentimen pengguna e-money pada twitter menggunakan algoritma c4.5dan naïve bayes ", *Jurnal Informatika dan Sistem Informasi*, Vol 1, No. 03, November 2020.
- [8] Kurniawan, Y. I. (2018). Perbandingan Algoritma Naive Bayes dan C.45 dalam Klasifikasi Data Mining. *Jurnal Teknologi Informasi Dan Ilmu Komputer*, 5(4), 455–464. <https://doi.org/10.25126/jtiik.20185480>.