

Analisis Sentimen Vaksinasi Booster Berdasarkan Twitter Menggunakan Algoritma Naïve Bayes dan K-NN

Afid Rozaqi¹, Agung Triayudi^{1,*}, Rima Tamara Aldisa²

¹Sistem Informasi, Fakultas Teknologi Komunikasi dan Informatika, Universitas Nasional, Jakarta, Indonesia

²Informatika, Fakultas Teknologi Komunikasi dan Informatika, Universitas Nasional, Jakarta, Indonesia

Email: ¹afid.rozaqi@gmail.com, ^{2,*}agungtriayudi@civitas.unas.ac.id, ³rimatamaraa@gmail.com

Email Penulis Korespondensi: agungtriayudi@civitas.unas.ac.id

Submitted: 26/09/2022; Accepted: 30/09/2022; Published: 30/09/2022

Abstrak-COVID-19 atau virus corona telah menyebar keseluruh dunia, salah satunya negara Indonesia. Telah banyak permasalahan akibat virus ini selama 2 tahun di Indonesia, dan telah dilakukan berbagai upaya dan kebijakan oleh pemerintah untuk mengendalikan dampak yang tidak menjadi lebih buruk oleh virus corona ini, upaya tersebut adalah tindakan vaksinasi terhadap masyarakat Indonesia, dan di awal tahun 2022 ini pemerintah memulai program baru yaitu vaksinasi booster. Banyak tanggapan masyarakat yang pro maupun kontra terhadap program tersebut di Twitter. Tujuan penelitian ini dilakukan untuk mengetahui sentimen warga Indonesia terhadap vaksinasi booster di Indonesia. Data yang didapat sebanyak 2000 tweet yang didapat dari kata kunci “vaksin booster” pada Twitter. Kemudian data dibagi menjadi data latih (training) dan data uji (testing) kemudian dijadikan tiga porsi yang berbeda yaitu 60/40, 70/30, dan 80/20. Hasil pengujian yaitu bahwa performa yang terbaik terdapat pada pengujian porsi 80% data latih 20% data uji menggunakan algoritma K-NN, pengujian tersebut menghasilkan hasil nilai paling tinggi yaitu akurasi 78.62% dan AUC 0.845 dan dikategorikan sebagai good classification. Dengan hasil tersebut menunjukkan bahwa model algoritma K-NN dengan porsi 80% data latih (training) adalah yang terbaik dalam klasifikasi analisis sentimen vaksinasi booster. Hasil sentimen pada data uji yaitu, positif berjumlah 303 data tweet dan sentimen negatif berjumlah 93 data tweet. Hasil sentimen positif yang lebih banyak menunjukkan bahwa vaksinasi booster di Indonesia dapat diterima dan mendapatkan banyak dukungan dari masyarakat Indonesia pada media sosial Twitter.

Kata kunci: Virus Corona; Vaksinasi Booster; Sentimen; Naïve Bayes; K-NN.

Abstract-Covid-19 or the corona virus has spread throughout the world, one of which is Indonesia. There have been many problems due to this virus for 2 years in Indonesia, and various efforts and policies have been made by the government to control the impact that does not become worse by this corona virus, these efforts are vaccination actions against the Indonesian people, and in early 2022 the government started a new program, namely booster vaccination. Many people are pro and contra to the program on social media Twitter. This study was conducted with the aim of knowing the sentiment of Indonesians towards booster vaccination in Indonesia. The data obtained as many as 2000 tweets obtained from the keyword "booster vaccine" on Twitter. Then the data is divided into training data and test data (training) then made into three different portions, namely 60/40, 70/30, and 80/20. The test results are that the best performance is found in testing a portion of 80% of the training data 20% of the test data using the K-NN algorithm, the test produced the highest value results, namely 78.62% accuracy and AUC 0.845 and categorized as good classification. The results show that the K-NN algorithm model with an 80% portion of training data is the best in the classification of booster vaccination sentiment analysis. The sentiment results in the test data were positive with 303 tweet data and negative sentiment totaled 93 tweet data. The results of more positive sentiments show that booster vaccinations in Indonesia are acceptable and get a lot of support from the Indonesian people on social media Twitter.

Keywords: Coronavirus; Booster Vaccination; Sentiment; Naïve Bayes; K-NN.

1. PENDAHULUAN

Pada akhir tahun 2019 di Wuhan, China telah terjadi kemunculan pertama suatu virus bernama corona atau *COVID-19*. Virus corona sangatlah berbahaya yang dimana penyebarannya sangat cepat, dan sudah hampir menyeluruh ke seluruh penjuru dunia. *WHO* menginformasikan bahwa sudah banyak negara-negara yang melaporkan bahwa negaranya sudah terjangkit virus corona atau *COVID-19* [1]. Indonesia adalah salah satu negara yang juga terdampak wabah virus corona. Virus ini pertama kali di Indonesia pada Maret 2020, yang pertama kali terjangkit adalah warga Depok, Jawa Barat [2]. Sejak saat kejadian itu, diketahui semakin banyaknya yang mulai terjangkit virus corona ini hingga menyeluruh ke seluruh daerah Indonesia. Pandemi yang semakin meluas di Indonesia mengakibatkan permasalahan-permasalahan pada berbagai bidang sektor, seperti pendidikan, ekonomi, kesehatan. Dengan kondisi negara yang mengalami pandemi tersebut, maka pemerintah Indonesia melakukan upaya-upaya dan kebijakan untuk mencegah dan mengurangi persebaran virus corona ini agar tidak banyak lagi dampak negatif yang akan terjadi pada masyarakat Indonesia. Vaksinasi adalah salah satu cara upaya pemerintah menekan penyebaran virus corona ini, dari tahun 2020 sampai 2021 sudah dilakukannya 2 tahap atau dosis vaksin yang mbahwa akan dilakukannya program vaksinasi *COVID-19* dosis ketiga (*booster*). Sudah tiga bulan program vaksinasi dosis ketiga dilaksanakannya, banyak berbagai tanggapan dan opini dari masyarakat Indonesia, mulai dari respon yang mendukung maupun yang bertentangan ataupun menolak. Dalam menganalisis pendapat atau opini masyarakat yang terdapat dimedia sosial yaitu *Twitter*, bisa dilakukannya dengan berbagai metode analisis yaitu di antaranya metode analisis sentimen.

Analisis sentimen adalah cabang dari data mining yang digunakan untuk menganalisis data tekstual berupa opini untuk menghasilkan informasi dengan nilai positif dan negatif [3]. Analisis sentimen adalah studi untuk

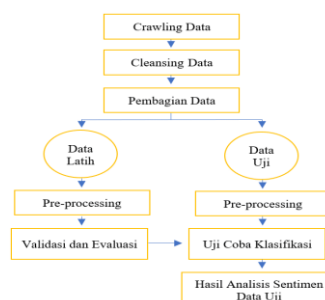
menganalisis positif atau pendapat negatif yang diungkapkan oleh pengguna pada media sosial seperti *Facebook*, *Twitter*, atau dalam data sosial[4]. *Data mining* merupakan suatu proses yang bertujuan untuk mencari informasi atau pola pada data yang sudah dipilih dengan menggunakan metode tertentu [5]. Terdapat beberapa model atau fungsi pada data mining yaitu *association rule mining*, *classification*, *feature selection*, dan *clustering* [6]. Klasifikasi memiliki algoritma-algoritma yang bisa digunakan atau diimplementasikan dalam suatu data. Dari algoritma-algoritma itu tentunya terdapat salah satu yang mempunyai akurasi tertinggi dan akurasi klasifikasi dapat mempengaruhi kinerja klasifikasi[7]. Akurasi algoritma dapat dihitung secara manual berdasarkan *Confusion Matrix* atau dapat dilihat menggunakan perangkat lunak yaitu *Rapidminer*[8]. Beberapa penelitian mengenai analisis sentimen sudah dilakukan sebelumnya. Seperti pada penelitian yang sudah dilakukan oleh Slamet Harry Ramadhani mengenai analisis sentimen vaksinasi *Astra Zeneca* pada media sosial *Twitter* dengan menggunakan algoritma *Naïve Bayes* dan *K-NN*. Hasil dari penelitian ini mendapatkan tingkat akurasi yang berbeda, yaitu pada pengujian *Naïve Bayes* mempunyai akurasi bernilai 90.71%, pada pengujian *K-NN* menghasilkan akurasi bernilai 74.78%[9]. Penelitian yang dilakukan oleh Muhammad Syarifuddin mengenai analisis sentimen mengenai *COVID-19* pada website *twitter* memakai algoritma *Naïve Bayes* dan *K-NN*. Penelitian ini menghasilkan nilai yang tinggi pada algoritma *Naïve Bayes* yaitu nilai *accuracy* sebesar 63.21% dan *precision* sebesar 66.40% dibandingkan algoritma *K-NN* pada nilai *accuracy* sebesar 58.10%. Didaptkan opini masyarakat pada *twitter* yang memiliki kecenderungan positif yang berdasarkan dari jumlah opini yang positif yaitu sebesar 610 dibandingkan dengan opini negatif sebesar 488[10].

Penelitian yang dilakukan oleh Achmad Bayhaqy tentang analisis sentimen E-commerce di Bukalapak dan Tokopedia pada media sosial *twitter*. Penelitian menghasilkan bahwa algoritma *Naïve Bayes* mempunyai nilai tertinggi dibandingkan dengan metode algoritma lainnya, yaitu nilai akurasi sebesar 77%, presisi 88,50%, dan *recall* 64%[11]. Selanjutnya penelitian yang dilakukan oleh Hanif Sudira tentang analisis sentimen terhadap kepuasan pelanggan terhadap digital layanan pembayaran (*Go-Pay*, *Ovo*, dan *LinkAja*) di Indonesia pada media sosial *instagram*. Hasil penelitian ini dengan 3800 data latih dan 200 data uji dengan menggunakan 20 k-fold cross validation menunjukkan bahwa *GO-PAY* memiliki semua jenis pelanggan yang puas, sedangkan *OVO* dan *LinkAja* mempunyai lebih banyak sentimen netral. Algoritma klasifikasi yang paling akurat diimplementasikan dalam penelitian ini adalah algoritma pengklasifikasi *K-NN*, meningkat hingga 90%. Hasil analisis sentimen akhir dari penelitian ini berkontribusi pada kondisi kepuasan pelanggan saat ini terhadap *Go-Pay*, *Ovo*, dan *LinkAja* sebagai penyedia layanan pembayaran digital di Indonesia dan mungkin masukan kepada pemangku kepentingan terkait[12]. Banyaknya penelitian terdahulu seperti diatas, membuktikan bahwa dapat dilakukannya analisis sentimen menggunakan pendekatan *machine learning* untuk mengetahui sentimen masyarakat terhadap masalah yang sedang terjadi atau topik yang sedang dibicarakan di media sosial. Berdasarkan latar belakang diatas, didapatkannya suatu permasalahan tentang “analisis sentimen opini masyarakat tentang Vaksinasi *Booster* di media sosial *twitter*”. Data tweet yang dipakai berjumlah 2000 *tweet*, dengan kata kunci ”vaksin booster” Tujuan dilakukannya penelitian ini adalah dapat mengelompokkan opini masyarakat menjadi dua kelas yaitu negatif dan positif menggunakan algoritma *Naïve Bayes* dan *K-NN* sehingga bisa mengetahui opini masyarakat terhadap vaksinasi dosis ketiga (*booster*), penelitian ini juga bertujuan menghasilkan nilai *accuracy* dan hasil AUC untuk mengetahui algoritma terbaik dalam mengklasifikasi sentimen vaksinasi booster. *Naïve Bayes* merupakan algoritma yang sangat baik digunakan pada data yang jumlahnya banyak dan bisa *missing value* atau data yang kosong dengan memakai perhitungan probabilitas dan statistik yang memprediksi suatu data berdasarkan pada pengujian yang dilakukan sebelumnya[13].

Salah satu algoritma pada model klasifikasi yang lebih baik dibandingkan dengan algoritma pengklasifikasi lainnya yaitu *Naïve Bayes* [14]. *K-NN* adalah metode supervised yang artinya memerlukan data latih (*training*) untuk proses mengklasifikasi suatu objek dengan jarak terdekat. Perhitungan pada metode *K-NN* yaitu data yang sudah melewati pra proses akan dipakai menjadi data *training* untuk data yang ingin diujikan, dengan mengukur kedekatan yang ada[15].

2. METODOLOGI PENELITIAN

Berikut gambar 1 merupakan tahapan dari penelitian yang dilakukan.



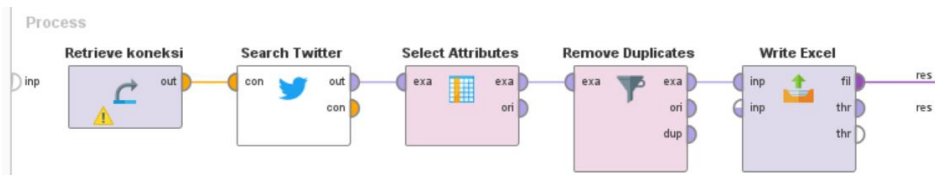
Gambar 1. Tahapan penelitian

Pada gambar 1, dijelaskannya tahapan-tahapan proses penelitian ini. Tahapan pertama yaitu dari pengumpulan data (*crawling data*), pelabelan (*labelling*) kemudian *pre-processing*, setelah itu dilakukan validasi dan evaluasi yang bertujuan untuk pengujian algoritma Naive Bayes dan K-NN, pembelajaran pada algoritma yang telah dilakukan sebelumnya diuji cobakan pada data uji yang akan menghasilkan analisis sentimen pada data uji.

3. HASIL DAN PEMBAHASAN

3.1 Crawling Data

Crawling data adalah proses awal dalam penelitian ini, proses *crawling data* ini menggunakan tools *Rapid Miner* dengan cara menghubungkannya pada *API-Twitter*, kemudian mencari data *tweet* dengan *keyword* “vaksin booster”. Pada penelitian ini dilakukannya dua pengujian berdasarkan jumlah data yang berbeda. Data yang sudah terkumpul kemudian disimpan ke dalam file *excel*. Proses *crawling data* dapat dilihat pada gambar 2. Kemudian dilakukannya proses *cleaning data* untuk menghilangkan karakter-karakter tertentu seperti karakter RT, @, dan URL pada teks data *tweet*. Proses *clensing data* dapat dilihat pada gambar 3.



Gambar 2. Crawling Data menggunakan Rapid Miner



Gambar 3. Proses Cleaning Data

3.2 Rancangan Skenario

Dalam penelitian ini, pengujian performa pada dua metode algoritma akan dilakukan. Untuk itu penulis berencana membuat skenario dengan porsi yang berbeda pada data latih (*training*) dan data uji (*testing*). Oleh karena itu, dengan skenario ini berharap dapat menunjukkan bagaimana performa masing-masing metode algoritma yang diuji berdasarkan porsi yang berbeda-beda dan dapat mengetahui mana yang terbaik untuk mengklasifikasi analisis sentimen vaksinasi *booster*. Tabel 1 menunjukkan skenario yang dibuat. Proporsi yang sering digunakan oleh peneliti yaitu 80/20, 70/30, atau 60/40[16].

Tabel 1. Tabel Rencana Simulasi

PORSI DATA				
	Latih (<i>Training</i>)	Uji (<i>Testing</i>)	Latih (<i>Training</i>)	Uji (<i>Testing</i>)
60/40	60%	40%	1200	800
70/30	70%	30%	1400	600
80/20	80%	20%	1600	400

3.3 Pre-processing

Setelah tahapan *crawling data*, didapatkan sekumpulan data (*dataset*) yang banyak yang masih dalam berbentuk *unstructured data* yaitu data yang tidak terstruktur dan belum siap dilakukannya analisis lebih lanjut maka dataset harus terlebih dahulu melalui tahapan proses *text pre-processing* yang bertujuan untuk menghilangkan atau mengatasi *noisy* pada data agar hasil dari perhitungan menjadi lebih baik. Alur pada *text pre-processing* yaitu, *case folding*, *filtering*, *tokenizing*, dan *stopword removal*[17].



Gambar 4. Pre-processing pada tools Rapid Miner.

3.4 Validasi dan Evaluasi

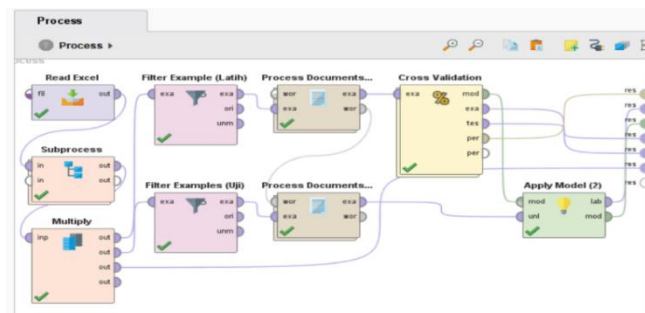
Kemudian, pada tahap validasi digunakannya operator *Cross Validation* yang di dalamnya terdapat proses-proses yang dapat digunakan untuk melakukan pengklasifikasian analisis sentimen. Pada tahapan ini, dilakukannya juga proses pelatihan dan implementasi algoritma *Naïve Bayes* dan *K-NN* dan juga dilakukannya proses-proses yang menghasilkan prediksi kelas sentimen atau label pada data latih. Pada gambar 5 dapat dilihat proses-proses yang terdapat pada operator validasi tersebut.



Gambar 5. Operator Validasi dan Evaluasi

Dalam pengujian terhadap data testing, perlu dilakukannya evaluasi terhadap hasil prediksi. Hasil pengujian *testing* menggunakan algoritma *Naïve Bayes* dan *K-NN* divisualisasikan pada *Confusion Matrix* untuk melihat nilai akurasi (*accuracy*). Pada proses ini juga dilakukannya pengujian pada dua algoritma yang menghasilkan nilai AUC. Terdapat panduan *range* untuk mengklasifikasikan keakuratan menggunakan *AUC* yaitu [18]: 0.50-0.60 (*Failure*), 0.70-0.80 (*Fair Classification*), 0.60-0,70 (*Poor Classification*), 0.80-0.90 (*Good Classification*), dan 0.90-1.00 (*Excellent Classification*).

Pada penelitian ini, proses analisis sentimen dilakukan menggunakan tools *RapidMiner*. Keseluruhan proses analisis sentimen dapat dilihat pada gambar di bawah ini.



Gambar 6. Proses keseluruhan analisis sentimen

Pada gambar diatas, terdapat operator *Read Excel* untuk *import* data hasil dari *crawling* data, operator *Subprocess* digunakan untuk *cleansing* data, operator *Multiply* digunakan untuk menduplikat data untuk kebutuhan proses *training* dan *testing*, operator *Filter Example* digunakan untuk memfilter data menjadi data latih atau data uji, operator *Process Documents* digunakan untuk proses *Preprocessing* Data, operator *Cross Validation* digunakan untuk proses validasi yang di dalamnya terdapat terdapat proses pelatihan dan penerapan algoritma data *mining*, pada operator *Apply Model (2)* untuk mengimplementasikan model algoritma yang telah dilatih sebelumnya menggunakan data latih dan bertujuan untuk melakukan klasifikasi analisis sentimen pada data uji yang belum mempunyai label.

3.5 Crawling Data

Dari proses *crawling* data menggunakan tools *RapidMiner* pada gambar 1 berhasil mendapatkan data berjumlah 2000 data yang didapat dari kata kunci “vaksin booster”. Data tersebut kemudian dibagi menjadi tiga porsi data yaitu 60% data *training* atau latih sebanyak 1200 data kemudian 70% data *training* sebanyak 1400 data dan 80% data *training* sebanyak 1600 data kemudian. Data dibagi juga menjadi 1000 *tweet* dan 2000 *tweet*. Pada 1000 *tweet*, data *testing* berjumlah 800 *tweet* dan data uji berjumlah 200 *tweet*. Kemudian pada 2000 data, data *testing* berjumlah 1600 *tweet* dan data uji berjumlah 400 *tweet*. Selanjutnya dilakukannya *labelling* secara manual pada data latih dengan kategori positif dan netral seperti pada gambar 7.

1	Text	Sentiment
2	ayo vaksin booster, lindungi diri kita dan keluarga dari covid-19 https://t.co/y3BvNBkkCB	POSITIF
3	ayo tuntas vaksin 1,2 hingga booster https://t.co/yqcEVh5rUZ	POSITIF
4	@myputun Vaksin booster di perlukan untuk menjaga tubuh dari varian covid-19. lagi pula vaksin juga gratis. https://t.co/GxM31ANkvd	POSITIF
5	Padahal takut bgt mau vaksin booster tp apakah daya jadi syarat wisuda??	NEGATIF
6	Antara nyesel ato ngga sok'an vaksin booster, badan sakit semua njrrr???	NEGATIF
7	gilak efek vaksin booster ke gw gini amat. Delay 1 hari dari tanggal vaksin. Asli macem habis dihantam mobil aja ini badan.demi osha oshi doang ini.aslinya males gw vaksin2 m.	NEGATIF
8	Ayo yang belum vaksinasi booster, vaksin segera ya! https://t.co/JcMcbubQkL	POSITIF

Gambar 7. Contoh Data Latih Yang Telah Diberi Label

Tabel 2. Hasil Sentimen Data *Training*

		POSITIF	NEGATIF
60%	1200 Data <i>Training</i>	794	406
70%	1400 Data <i>Training</i>	941	459
80%	1600 Data <i>Training</i>	1109	491

3.6 Pre-processing

Berikut gambar 8 data latih sebelum pre-processing dan gambar 8 setelah pre-processing.

Row No.	Sentiment	Text
1	POSITIF	ayo vaksin booster, lindungi diri kita dan keluarga dari covid-19
2	POSITIF	ayo tuntas vaksin 1,2 hingga booster
3	POSITIF	@myputun Vaksin booster di perlukan untuk menjaga tubuh dari varian covid-19. lagi pula vaksin juga gratis.
4	NEGATIF	Padahal takut bgt mau vaksin booster tp apakah daya jadi syarat wisuda??
5	NEGATIF	Antara nyesel ato ngga sok'an vaksin booster, badan sakit semua njrrr????
6	NEGATIF	gilak efek vaksin booster ke gw gini amat. Delay 1 hari dari tanggal vaksin. Asli macem habis dihantam mobil aja ini badan.
7	POSITIF	Ayo yang belum vaksinasi booster, vaksin segera yal
8	NEGATIF	@fadlizon Booster cuma membuat orang lebih rentan, membahayakan rakyat dan memperburuk & memperpanjang pandemi.
9	POSITIF	Nah lo percuma vaksin booster masih kena covid siapa yang tanggung jawab semua ini

Gambar 8. Data Latih Sebelum *Pre-Processing*

Row No.	Sentiment	text	aatra	abis
1	POSITIF	ayo vaksin booster lindungi keluarga covid	0	0
2	POSITIF	tuntas vaksin booster	0	0
3	POSITIF	myputun vaksin booster perlukan menjaga tubuh varian covid vaksin gratis	0	0
4	NEGATIF	takut vaksin booster apakah daya syarat wisuda	0	0
5	NEGATIF	nyesel ngga vaksin booster badan sakit njrr	0	0
6	NEGATIF	gilak efek vaksin booster gini delay tanggal vaksin asli macem habis dihantam mobil badan osha oshi doang aslinya males vaksin mulu	0	0
7	POSITIF	vaksinasi booster vaksin	0	0
8	NEGATIF	fadlizon booster orang rentan membahayakan rakyat memperburuk memperpanjang pandemi baca data diajukan bukti gugatan vaksin ...	0	0
9	POSITIF	vaksin booster kena covid tanggung	0	0

Gambar 9. Hasil Data Latih Setelah *Pre-Proceesing*

Setelah tahap data *cleaning*, selanjutnya dilakukannya *pre-processing* pada data latih yang bertujuan untuk menghasilkan sebuah teks menjadi data yang mudah diproses oleh sistem. Tahap *pre-processing* dibutuhkan sebelum melakukan evaluasi atau pemodelan algoritma. Pada tahap ini dilakukan *tokenize* untuk memecah kalimat pada teks menjadi beberapa kata-kata, lalu *transform case* untuk mengubah teks menjadi huruf kecil semua, filter tokens untuk menghilangkan kata yang jumlah hurufnya tidak memenuhi *range* yang ditetapkan, dan yang terakhir filter *stopwords* untuk menghilangkan kata-kata yang tidak memiliki arti. Perubahan pada teks data latih sebelum dan sesudah *pre-processing* dapat dilihat pada gambar 8 dan gambar 9.

3.7 Hasil Validasi dan Evaluasi

Setelah melewati proses *pre-processing*, data latih (*training*) digunakan sebagai pembelajaran atau melatih algoritma untuk proses klasifikasi data uji dengan menggunakan algoritma *Naïve Bayes* dan *K-NN*. Pengujian dengan dua algoritma menggunakan operator *Cross Validation* yang akan menghasilkan nilai *accuracy* dan *AUC*.

Tabel 3. Contoh Hasil Evaluasi Pada Dua Jenis Jumlah Data Latih (*Training*)

JUMLAH DATA TRAINING	<i>Naïve Bayes</i>		<i>K-NN</i>	
	<i>Accuracy</i>	<i>AUC</i>	<i>Accuracy</i>	<i>AUC</i>
60% (1200 Data <i>Training</i>)	67.50%	0.553%	75.33%	0.810%
70% (1400 Data <i>Training</i>)	70.50%	0.590%	76.07%	0.828%
80% (1600 Data <i>Training</i>)	72.38%	0.535%	78.62%	0.845%

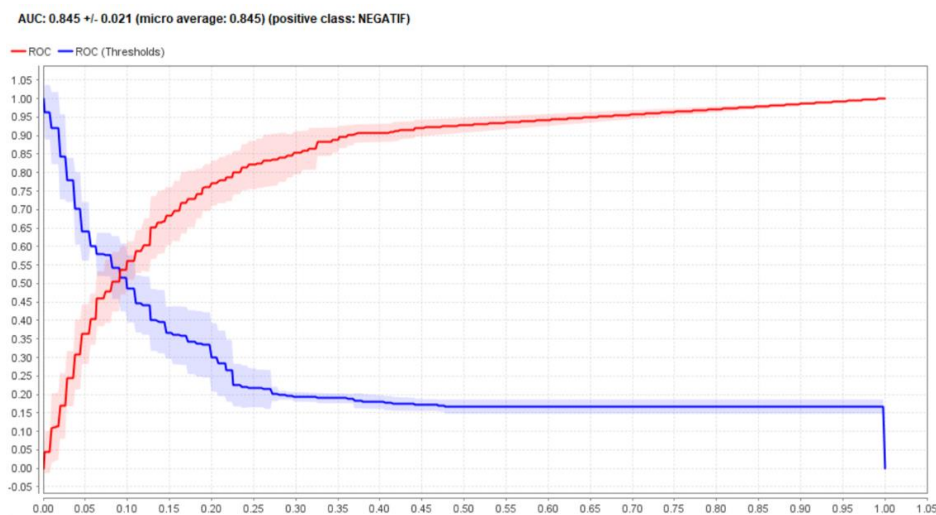
Dari tabel 3 diatas, dari pengujian 60% data *training* menggunakan algoritma *Naïve Bayes* menghasilkan nilai *accuracy* yang lebih besar, yaitu 75.33% dan nilai *AUC* yang lebih besar juga yaitu 0.810%, dengan nilai tersebut algoritma *K-NN* lebih baik dan dikategorikan *good classification*. Lalu pada pengujian 70% data *training*, menghasilkan nilai *accuracy* 70.50% dan nilai *AUC* sebesar 0.590% dengan menggunakan algoritma *Naive Bayes*, pada algoritma *K-NN* menghasilkan nilai yang lebih besar yaitu *accuracy* sebesar 76.07% dan nilai *AUC* sebesar 0.828% yang berarti dikategorikan sebagai *good classification*. Pengujian terakhir yang dilakukan menggunakan 80% data *training* menghasilkan nilai *accuracy* 72.38% dan *AUC* 0.535% menggunakan algoritma *Naive Bayes*, kemudian menggunakan algoritma *K-NN* menghasilkan nilai yang lebih besar yaitu *accuracy* sebesar 78.62% dan *AUC* sebesar 0.845% yang berarti dikategorikan dalam *good classification*.

Berdasarkan pada pengujian tiga jumlah data training yang berbeda menggunakan dua algoritma, Pada gambar 10 dan 11, algoritma K-NN selalu menunjukkan nilai *accuracy* dan AUC yang selalu lebih tinggi dibandingkan algoritma *Naive Bayes* disetiap pengujiannya dan pada pengujian 80% data *training* merupakan yang paling baik dalam menghasilkan nilai *accuracy* dan AUC yang paling tinggi dibandingkan pengujian menggunakan 60% dan 70% data *training*. Nilai *accuracy* dan AUC terbaik terdapat pada pengujian 80% data *training* menggunakan algoritma K-NN yaitu nilai *accuracy* sebesar 76.82% dan AUC sebesar 0.845%, dengan begitu dikategorikan *good classification*. Algoritma K-NN dengan porsi 80% data *training* dan 20% data uji, merupakan menjadi model terbaik untuk klasifikasi analisis sentimen vaksinasi *booster*.

accuracy: 78.62% +/- 2.08% (micro average: 78.62%)

	true POSITIF	true NEGATIF	class precision
pred. POSITIF	1003	236	80.95%
pred. NEGATIF	106	255	70.64%
class recall	90.44%	51.93%	

Gambar 10. Accuracy K-NN 80% Data Training



Gambar 11. AUC K-NN 80% Data Training

Hasil nilai akurasi pada confusion matrix juga dapat dihitung secara manual seperti yang dibawah ini:

$$\text{Accuracy} : \frac{\text{True Negative} + \text{True Positive}}{\text{Total Data Matrix}} \times 100\% \quad (19)$$

$$\text{Accuracy} : \frac{1003+255}{1003+255+106+236} \times 100\%$$

$$= \frac{1258}{1600} \times 100\%$$

$$= 0,78625 \times 100\% = 78,62\%$$

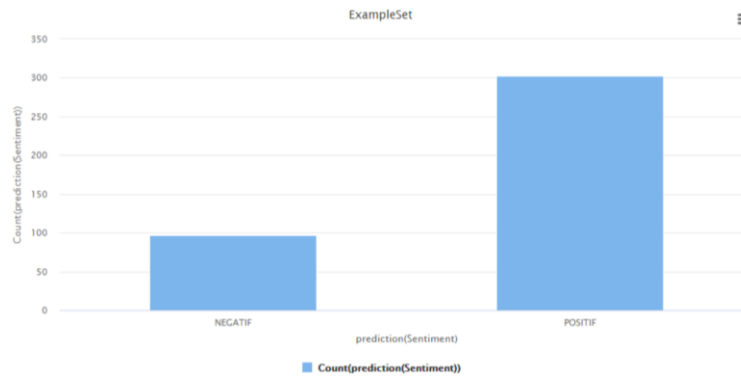
Akurasi menunjukkan seberapa akurat metode dalam mengklasifikasikan data dengan baik. Dengan begitu, bisa mengetahui sejauh mana akurat nilai akurasi pada kinerja sistem untuk mengetahui informasi yang dibutuhkan pengguna (*user*) menggunakan hasil yang disediakan oleh sistem[20].

3.8 Hasil Analisis Sentimen

Setelah selesai melakukan proses *crawling data*, *cleansing data*, *pre-processing*, dan evaluasi, maka akan didapatkan hasil analisis sentimen pada data uji.

Tabel 4. Hasil analisis sentimen data uji

	K-NN	
	POSITIF	NEGATIF
20% (400 data uji)	303	97



Gambar 12. Diagram Hasil Sentimen Data Uji

Tabel di atas dihasilkan dari operator *apply model* (2) yang memperlihatkan hasil analisis sentimen pada pengujian 20% data uji menggunakan model algoritma K-NN, menghasilkan sentimen positif sebanyak 303 data dan sentimen negatif sebanyak 97 negatif. Pada gambar 12 adalah result *apply model* (2) yang memperlihatkan beberapa data uji yang telah diprediksi sentimennya berdasarkan besaran nilai *confidence* yang mendekati nilai 1.

Row No.	prediction(S...	confidence(...	confidence(...	text	aamin	aatra	ab
1	NEGATIF	0	1	abis vaksin booster bawaannya ngantuk mulu astaga	0	0	0
2	POSITIF	1	0	vaksin booster pastinkirlebikuat	0	0	0
3	NEGATIF	0.396	0.604	semenjak vaksin berat badanku booster	0	0	0

Gambar 13. Prediksi Sentimen Data Uji.

4. KESIMPULAN

Hasil pengujian pada skenario tiga jenis porsi jumlah data *training* yang berbeda dengan menggunakan algoritma *Naive Bayes* dan *K-NN* menunjukkan bahwa performa yang terbaik terdapat pada pengujian porsi 80% data latih (*training*) 20% data uji (*testing*) dengan menggunakan algoritma *K-NN*, pengujian tersebut menghasilkan hasil nilai paling tinggi yaitu akurasi (*accuracy*) sebesar 78.62% dan *AUC* sebesar 0.845 yang artinya dapat dikategorikan dalam *good classification*. Dengan hasil tersebut menunjukkan bahwa model algoritma *K-NN* dengan porsi 80% data latih (*training*) 20% data uji (*testing*) adalah yang terbaik dalam klasifikasi analisis sentimen *vaksinasi booster*. Pengujian dengan porsi 80% data latih (*training*) 20% data uji (*testing*) dengan menggunakan algoritma *K-NN* juga berhasil memprediksi sentimen pada data uji, yaitu sentimen positif berjumlah 303 data tweet dan sentimen negatif berjumlah 93 data *tweet*. Hasil sentimen positif yang lebih banyak dibandingkan sentimen negatif tersebut dapat disimpulkan bahwa *vaksinasi booster* di Indonesia dapat diterima dan mendapatkan banyak dukungan dari masyarakat Indonesia pada media sosial *Twitter*.

REFERENCES

- [1] A. F. Watratan, A. P. B, D. Moeis, S. Informasi, and S. P. Makassar, (2020) "Journal of applied computer science and technology (JACOST) Implementasi Algoritma Naive Bayes Untuk Memprediksi Tingkat Penyebaran Covid-19 Di Indonesia," vol. 1, no. 1, pp. 7–14, 2020.
- [2] A. Susilo *et al.*, (2020) "Coronavirus Disease 2019: Tinjauan Literatur Terkini," *J. Penyakit Dalam Indones.*, vol. 7, no. 1, p. 45, 2020, doi: 10.7454/jpdi.v7i1.415.
- [3] D. Ramadhan and E. B. Setiawan, (2019) "Analisis Sentimen Program Acara di SCTV pada Twitter Menggunakan Metode Naive Bayes dan Support Vector Machine," *Telkomuniversity.Ac.Id*, vol. 6, no. 2, pp. 9736–9743, 2019, [Online]. Available: <https://openlibrarypublications.telkomuniversity.ac.id/index.php/engineering/article/view/10708>.
- [4] S. H. Lee, Y. W. Cho, E. T. Im, and G. Y. Gim, (2019) "A Study on Customer Satisfaction Analysis of Public Institutions using Social Textmining," *Proc. - 20th IEEE/ACIS Int. Conf. Softw. Eng. Artif. Intell. Netw. Parallel/Distributed Comput. SNPD 2019*, pp. 385–394, 2019, doi: 10.1109/SNPD.2019.8935791.
- [5] I. Hidayanti, T. B. Kurniawan, and A. Afriyudi, (2020) "Perbandingan Dan Analisis Metode Klasifikasi Untuk Menentukan Konsentrasi Jurusan," *J. Ilm. Inform. Glob.*, vol. 11, no. 1, pp. 16–21, 2020, doi: 10.36982/jig.v11i1.1067.
- [6] I. Lishania, R. Goejantoro, and Y. N. Nasution, (2019). "Perbandingan Klasifikasi Metode Naive Bayes dan Metode Decision Tree Algoritma (J48) pada Pasien Penderita Penyakit Stroke di RSUD Abdul Wahab Sjahranie Samarinda," *J. Eksponensial*, vol. 10, no. 2, pp. 135–142, 2019, [Online]. Available: <http://jurnal.fmipa.unmul.ac.id/index.php/exponensial/article/view/571>.
- [7] U. Riyanto, (2019) "Analisis Perbandingan Algoritma Naive Bayes Dan Support Vector Machine Dalam Mengklasifikasikan Jumlah Pembaca Artikel Online," *JIKA (Jurnal Inform.*, vol. 2, no. 2, pp. 62–72, 2019, doi: 10.31000/v2i2.1521.
- [8] Ainurrohmah, (2021) "Akurasi Algoritma Klasifikasi pada Software Rapidminer dan Weka," *Prisma*, vol. 4, pp. 493–499, 2021, [Online]. Available: <https://journal.unnes.ac.id/sju/index.php/prisma/>.

- [9] S. H. Ramadhani and M. I. Wahyudin, (2022) “Analisis Sentimen Terhadap Vaksinasi Astra Zeneca pada Twitter Menggunakan Metode Naïve Bayes dan K-NN,” *J. JTik (Jurnal Teknol. Inf. dan Komunikasi)*, vol. 6, no. 4, pp. 526–534, 2022, doi: 10.35870/jtik.v6i4.530.
- [10] M. Syarifuddin, (2022) “Analisis Sentimen Opini Publik Mengenai Covid-19 Pada Twitter Menggunakan Metode Naïve Bayes Dan Knn,” *Inti Nusa Mandiri*, vol. 15, no. 1, pp. 23–28, 2020.
- [11] A. Bayhaqy, (2018) “Sentiment Analysis about E-Commerce from Tweets Using Decision Tree , K-Nearest Neighbor , and Naïve Bayes,” *2018 Int. Conf. Orange Technol.*, pp. 1–6.
- [12] H. Sudira, A. L. Diar, and Y. Ruldeviyani, (2019) “Instagram Sentiment Analysis with Naive Bayes and KNN: Exploring Customer Satisfaction of Digital Payment Services in Indonesia,” *2019 Int. Work. Big Data Inf. Secur. IWBIS 2019*, pp. 21–26, 2019, doi: 10.1109/IWBIS.2019.8935700.
- [13] P. Studi and M. Informatika, (2018) “Perbandingan data mining SVM dan NN untuk klasifikasi penyakit ginjal kronisS,” vol. 14, no. 1, pp. 1–6, 2018.
- [14] D. Xhemali, C. J. Hinde, and R. G. Stone, (2009) “Naive Bayes vs. Decision Trees vs. Neural Networks in the Classification of Training Web Pages,” *Int. J. Comput. Sci.*, vol. 4, no. 1, pp. 16–23, 2009, [Online]. Available: <http://cogprints.org/6708/>.
- [15] S. Nasional, T. Elektro, S. Informasi, and T. Informatika, “Studi Komparasi Metode Klasifikasi K-Nearest Neighbor dan Naïve Bayes dalam Mengidentifikasi Kepuasan Pelanggan Terhadap Produk,” pp. 333–338, 2022.
- [16] S. Raschka, (2018) “Model Evaluation, Model Selection, and Algorithm Selection in Machine Learning,” 2018, [Online]. Available: <http://arxiv.org/abs/1811.12808>.
- [17] E. A. Sukma, A. N. Hidayanto, A. I. Pandesenda, A. N. Yahya, P. Widharto, and U. Rahardja, (2017) “Sentiment Analysis of the New Indonesian Government Policy (Omnibus Law) on Social Media Twitter,” *Proc. - 2nd Int. Conf. Informatics, Multimedia, Cyber, Inf. Syst. ICIMCIS 2020*, no. July 2017, pp. 153–158, 2020, doi: 10.1109/ICIMCIS51567.2020.9354287.
- [18] D. K. Widiyati, M. Wati, and H. S. Pakpahan, (2018) “Penerapan Algoritma ID3 Decision Tree Pada Penentuan Penerima Program Bantuan Pemerintah Daerah di Kabupaten Kutai Kartanegara,” *J. Rekayasa Teknol. Inf.*, vol. 2, no. 2, p. 125, 2018, doi: 10.30872/jurti.v2i2.1864.
- [19] M. M. Ikram, (2021) “Keputusan Penggunaan Layanan GoFood Selama Masa Pandemi Covid-19,” *J. Ilm. Manaj. Kesatuan*, vol. 9, no. 2, pp. 71–80, 2021, doi: 10.37641/jimkes.v9i2.467.
- [20] F. Fitriana, E. Utami, and H. Al Fatta, (2021) “Analisis Sentimen Opini Terhadap Vaksin Covid - 19 pada Media Sosial Twitter Menggunakan Support Vector Machine dan Naive Bayes,” *J. Komtika (Komputasi dan Inform.)*, vol. 5, no. 1, pp. 19–25, 2021, doi: 10.31603/komtika.v5i1.5185.
- [21] Aldisa, R. T., & Maulana, P. (2022). Analisis Sentimen Opini Masyarakat Terhadap Vaksinasi Booster COVID-19 Dengan Perbandingan Metode Naive Bayes, Decision Tree dan SVM. *Building of Informatics, Technology and Science (BITS)*, 4(1), 106-109.