

Penerapan Algoritma CNN Untuk Mengetahui Sentimen Masyarakat Terhadap Kebijakan Vaksin Covid-19

Fany Alifian Irawan¹, Dwi Anindyani Rochmah²

^{1,2} Universitas Mercu Buana
Jl. Raya, RT.4/RW.1, Meruya Sel., Kec. Kembangan, Jakarta Barat, DKI. Jakarta, Indonesia

e-mail: ¹141518110049@student.mercubuana.ac.id, ²dwi.anindya@mercubuana.ac.id

Informasi Artikel

Diterima: 05-07-2022

Direvisi: 26-09-2022

Disetujui: 30-09-2022

Abstrak

Jejaring sosial Twitter merupakan wadah bagi netizen dari seluruh dunia untuk bertukar pendapat dan argumen, beragam topik diangkat oleh netizen terutama permasalahan yang sedang hangat diperbincangkan atau menjadi perdebatan di khalayak umum. Salah satu topik yang hangat dibicarakan netizen Indonesia yaitu mengenai Vaksin Covid-19 yang merupakan salah satu kebijakan pemerintah Indonesia dalam upaya menanggulangi *pandemic* Covid-19. Seperti kebijakan lainnya yang tak luput menimbulkan pro-kontra, kebijakan vaksin ini juga menjadi perbincangan pada jejaring Twitter. Atas dasar itu untuk mendapatkan informasi yang terdapat pada komentar netizen di jejaring sosial Twitter, maka diperlukan analisis sentimen dengan tujuan mengetahui sebagian respon masyarakat Indonesia mengenai kebijakan vaksin sehingga dapat menjadi bahan pertimbangan pihak terkait dalam mengevaluasi kebijakan sehingga menjadi lebih baik. Analisa sentimen dilakukan dengan mengambil data komentar Twitter seputar vaksin yang dibuat menjadi dataset dengan dua polaritas sentimen positif dan negatif dengan nilai masing-masing sentimen sebesar 650 data. Dataset digunakan untuk menganalisa sentimen serta digunakan pada tahap pengujian tingkat akurasi algoritma. Berdasarkan hasil pengujian, algoritma *Convolutional Neural Network* memperoleh rata-rata nilai akurasi sebesar 98.66%, dengan algoritma pembandingan yaitu *Naïve Bayes* yang memperoleh rata-rata nilai akurasi sebesar 94.66%. Hasil dari penelitian dapat disimpulkan bahwa kebijakan vaksinasi ini mendapatkan respon yang positif berdasarkan data komentar Twitter yang berjumlah 1424 baris, sebanyak 950 komentar berpolaritas positif dengan persentase 66.7% dan 33.3% sisanya sejumlah 474 komentar berpolaritas negatif. Selain itu berdasarkan data *Wordcloud* diketahui sebagian besar komentar bermuatan negatif berisi dengan kata-kata yang menyiratkan efek samping dari vaksinasi terutama jenis vaksinasi *booster*.

Kata Kunci: *Convolutianal Neural Network*, *Naïve Bayes*, Sentimen Analisis

Abstract

The social network Twitter is a forum for netizens from all over the world to exchange opinions and arguments, various topics are raised by netizens, especially issues that are hotly discussed or become debates in the general audience. One of the topics that Indonesian netizens are hotly discussed is about the Covid-19 Vaccine, which is one of the Indonesian government's policies in an effort to overcome the Covid-19 pandemic. Like other policies that have not spared from causing pros and cons, this vaccine policy has also become a topic of discussion on the Twitter network. On that basis, to obtain information contained in the comments of netizens on the Twitter social network, a sentiment analysis is needed with the aim of knowing some of the responses of the Indonesian people regarding vaccine policies so that they can be considered by related parties in evaluating policies so that they become better. Sentiment analysis was carried out by taking data on Twitter comments about the vaccine which was made into a dataset with two polarities of positive and negative sentiments with a sentiment value of 650 data each. Datasets are used to analyze sentiment and are used at the testing stage of the accuracy level of the algorithm. Based on the test results, the Convolutional Neural Network algorithm obtained an average accuracy value of 98.66%, with a comparison algorithm, namely Naïve Bayes, which obtained an average accuracy value of 94.66%. The results of the study can be concluded that this vaccination policy received a positive response based on Twitter comment data totaling 1424 lines, as many as 950 comments with a positive polarity with a percentage of 66.7% and the remaining 33.3% a



total of 474 comments with a negative pattern. In addition, based on Wordcloud data, it is known that most of the negatively charged comments contain with words that imply side effects of vaccination, especially the type of booster vaccination.

Keywords: *Convolutianal Neural Network, Naïve Bayes, Sentiment Analysis*

1. Pendahuluan

Pada Desember 2019, kasus pneumonia misterius pertama kali dilaporkan di Wuhan, Provinsi Hubei. Sumber penularan kasus ini masih belum diketahui pasti, tetapi kasus pertama dikaitkan dengan pasar ikan di Wuhan. Tanggal 18 Desember hingga 29 Desember 2019, terdapat lima pasien yang dirawat dengan *Acute Respiratory Distress Syndrome* (ARDS). Sejak 31 Desember 2019 hingga 3 Januari 2020 kasus ini meningkat pesat, ditandai dengan dilaporkannya sebanyak 44 kasus. Tidak sampai satu bulan, penyakit ini telah menyebar di berbagai provinsi lain di China, Thailand, Jepang, dan Korea Selatan (Susilo et al., 2020). Demam, batuk, dan sesak napas adalah gejala yang paling umum dari corona virus. Dampak untuk lansia, penyakit penyerta kronis dan riwayat perjalanan dapat menjadi faktor risiko penyebab kematian (Khan et al., 2020).

Indonesia termasuk negara yang terdampak sebaran virus Covid-19. Tingkat penyebaran virus ini cukup cepat sehingga *World Health Organization* (WHO) telah menetapkan wabah ini sebagai sebuah pandemi penyakit (Ayumi & Nurhaida, 2021). Sebagai bentuk ikhtiar mencegah dan menanggulangi wabah Covid-19 yang semakin meningkat, pemerintah Indonesia resmi menerapkan kebijakan vaksin untuk masyarakat yang pertama kali dilakukan pada hari Rabu 13 Januari 2021 berlokasi di Istana Negara. Vaksin yang digunakan berjenis *Sinovac* dengan peserta vaksinasi yaitu Presiden Joko Widodo, beberapa pejabat pemerintah, tokoh agama, serta beberapa perwakilan masyarakat sipil. Resminya kebijakan vaksinasi tersebut menuai beragam tanggapan di berbagai lini media sosial salah satunya Twitter. Berbagai komentar dari netizen (*citizen of the net*) bermunculan di jejaring sosial *Twitter*, yang turut menghiasi lini masa media sosial tersebut mulai dari komentar yang bersifat pro maupun kontra. Komentar yang diberikan netizen (*citizen of the net*) sangat beragam, beberapa bersifat positif, negatif, dan ada pula yang bersifat netral.

Menanggapi kondisi tersebut maka perlu dilakukan *opinion mining* atau bisa disebut sentimen analisis. Sentimen analisis dilakukan untuk mengetahui tingkat kepuasan masyarakat dan dapat digunakan sebagai bahan evaluasi

peningkatan pelayanan (Dwianto et al., 2021). Analisa ini penting dilakukan untuk mengetahui gambaran secara garis besar persepsi masyarakat dalam menanggapi kebijakan pemerintah ini, apakah memiliki kecenderungan positif atau kecenderungan negatif. Analisis sentimen melibatkan komputasi dalam memahami sentimen tersirat di dalam sebuah teks menggunakan metode *deep learning*. *Deep learning* memiliki beberapa varian diantaranya jaringan syaraf konvolusional, jaringan syaraf berulang, dan jaringan syaraf rekursif (Ihsan, 2021).

Pada penelitian ini dilakukan analisis sentimen masyarakat terhadap kebijakan vaksinasi Covid-19 berdasarkan komentar pada jejaring sosial Twitter menggunakan teknik *Sentiment Analysis* untuk mengolah dan mengekstrak data dari Twitter. Hasil *preprocessing* berupa token-token kata selanjutnya diklasifikasikan dengan pendekatan *deep learning* menggunakan algoritma *Convolution Neural Network* (CNN).

Untuk dapat membangun jaringan syaraf tiruan pada algoritma CNN maka diperlukan *library* Keras. Keras merupakan *library* jaringan syaraf tiruan tingkat tinggi yang ditulis dengan bahasa python dan mampu berjalan di atas *TensorFlow*. *Tensorflow* adalah *library* python yang digunakan untuk menghitung nilai sub-grafik dengan cara membuat kompartemen dengan neural network sehingga menghasilkan pelatihan data yang terdistribusi dengan baik (Nasichuddin et al., 2018). *Library* ini menyediakan fitur yang digunakan dengan fokus mempermudah pengembangan lebih dalam tentang *Deep Learning* (Santoso & Ariyanto, 2018). Model CNN telah memecahkan permasalahan dalam pemrosesan gambar dan saat ini peneliti-peneliti telah mengembangkan CNN untuk NLP seperti analisis sentimen, klasifikasi polaritas emosional, *text summary* (Rhanoui et al., 2019). CNN mampu memuat keseluruhan skala informasi klasifikasi objek tanpa kehilangan keakuratannya (Dhika et al., 2020).

Dibandingkan dengan metode klasik *machine learning* seperti SVM, *Logistic Regression* dan *Naïve Bayes*, *DoubleMax* CNN bekerja rata-rata 17 % lebih baik dalam melakukan klasifikasi sentimen data. Hal ini

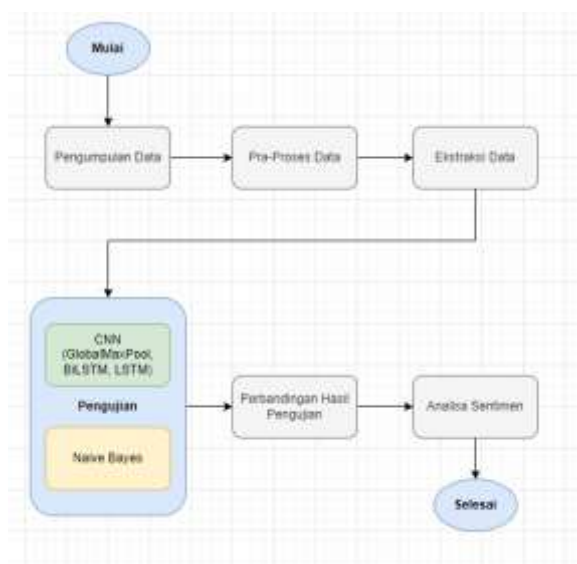
menunjukkan bahwa CNN dengan arsitektur yang lebih kompleks memiliki performa yang lebih baik dalam menganalisis sentiment (Qudsi et al., 2021).

Penelitian ini dilakukan untuk mengetahui sentimen masyarakat dalam menanggapi kebijakan yang dilakukan pemerintah menggunakan teknik sentimen analisis berdasarkan sampel komentar netizen (*citizen of the net*) pada jejaring twitter sehingga dapat dimanfaatkan untuk pertimbangan evaluasi terkait kebijakan yang telah dibuat maupun yang akan datang. Sampel komentar yang digunakan untuk mengetahui sentimen masyarakat terhadap kebijakan vaksin berjumlah 1424 baris komentar yang berisi tanggapan masyarakat seputar vaksin dengan nilai polaritas sentimen yang belum diketahui.

Sentimen analisis dilakukan dengan melakukan analisa terhadap media sosial twitter dimana dilakukan tahapan *crawling data* yang akan dijadikan dataset dan data sampel yang akan dianalisa nilai polaritasnya menggunakan teknik *sentiment analysis*.

2. Metode Penelitian

Pada penelitian tahapan yang digunakan secara keseluruhan dalam melakukan pengujian dan analisa sentimen dapat dilihat pada Gambar 1.



Gambar 1. Tahapan Penelitian

Berikut ini merupakan penjelasan dari Gambar yang merupakan alur tahapan penelitian untuk proses analisis sentimen

dengan melakukan uji coba perbandingan algoritma *Convolutional Neural Network* dan algoritma *Naïve Bayes* dalam mengolah *dataset* berdasarkan komentar Twitter Netizen (*citizen of the net*):

A. Pengumpulan Data

Dataset Dataset komentar masyarakat melalui media social Twitter dikumpulkan untuk analisis sentimen menggunakan API Twitter, komentar yang diambil merupakan komentar berbahasa Indonesia yang diambil pada rentang waktu Januari 2021 hingga sekarang, dengan menggunakan tool Rapid Miner Studio versi 9.8. Dalam melakukan pengumpulan data berpolaritas negatif pengambilan dilakukan beberapa kali dengan query yang digunakan bervariasi yaitu *vaksin nyeri*, *efek vaksin*, *vaksin sakit*, dan *vaksin alergi*. Sedangkan untuk mendapatkan komentar berpolaritas positif query yang digunakan yaitu *vaksin sehat*, *vaksin aman*, dan *sukses vaksin*. Dari proses pengambilan data komentar melalui API tersebut diperoleh 2148 data komentar, akan tetapi dilakukan proses seleksi dan *cleaning* data sehingga menjadi 1295 data. Dari data komentar yang diperoleh, terdapat 13 field yaitu *created-at*, *from-user*, *from-user-id*, *to-user*, *to-user-id*, *language*, *source*, *text*, *geo-location-latitude*, *geo-location-longitude*, *retweet-count*, *id*, dan *sentiment*. Untuk melakukan proses *training* dan *testing* data hanya menggunakan variabel *text* untuk dianalisa sentimennya.

B. Pra-Proses Data

Pra-Proses Data atau Preprocessing merupakan tahap yang harus dilakukan sebelum dilanjutkan ketahap pembelajaran atau *training data*. Tahap ini berisi proses pelabelan (*labeling data*) dan juga pembersihan data (*data cleaning*). Tahap pelabelan adalah proses memberikan nilai sentimen pada kalimat atau komentar berupa sentiment dengan polaritas positif dan negatif, proses ini dilakukan secara manual yang nantinya akan dipelajari oleh mesin dalam bentuk pembelajaran analisis sentimen.

Karena data yang dikumpulkan memiliki kata yang hilang atau tidak baku maka tahap selanjutnya adalah pembersihan data yaitu proses menyiapkan kata atau kalimat agar dapat diproses dalam pembelajaran data. Tahap ini sangat berguna untuk memudahkan atau memaksimalkan mesin dalam mengklasifikasikan kalimat yang ada kedalam polaritas sentiment. Beberapa tahap dalam pembersihan data yaitu *case folding*, *tokenize*, *stemming*, *remove punctuation*, dan *filtering* /

stopword removal. Proses pembersihan data menggunakan bantuan *library* pendukung dalam bahasa Python seperti NLTK, Sastrawi, Re, dan beberapa *library* pendukung lainnya. Berikut penjelasan mengenai tahap-tahap pembersihan data atau pra-proses data.

1. Penghapusan tanda baca (*Remove Punctuation*)

Penghapusan tanda baca atau *Remove Punctuation* merupakan proses penting dalam pra-proses data karena bila proses ini terlewat dapat mempengaruhi hasil pembelajaran data. Pada tahap ini dilakukan penghilangan tanda baca pada kalimat atau teks yang nantinya tidak diperlukan pada pembelajaran data dan mencegah perubahan makna karena adanya tanda baca. Berikut ditampilkan hasil dari tahap *remove user* pada Tabel 1. *remove punctuation* pada Tabel 2. dan Tabel 3. menampilkan tahap menghapus angka.

Tabel 1. Remove User dan remove RT

Teks	Hasil
rt @penyair_berdiri: vaksin covid-19 yang digunakan di indonesia sudah melalui uji klinis & terbukti aman. sudah melewati proses panjang	: vaksin covid-19 yang digunakan di indonesia sudah melalui uji klinis & terbukti aman. sudah melewati proses panjang

Tabel 2. Remove Punctuation

Teks	Hasil
: vaksin covid-19 yang digunakan di indonesia sudah melalui uji klinis & terbukti aman. sudah melewati proses panjang	vaksin covid19 yang digunakan di indonesia sudah melalui uji klinis bukti aman sudah melewati proses panjang

Tabel 3. Remove Angka

Teks	Hasil
vaksin covid19 yang digunakan di indonesia sudah melalui uji klinis terbukti aman sudah melewati proses panjang	vaksin covid yang digunakan di indonesia sudah melalui uji klinis bukti aman sudah melewati proses panjang

2. *Tokenize*

Tokenisasi adalah proses memotong sebuah kalimat atau teks menjadi bagian kecil atau kata yang terpisah yang disebut token. Tokenisasi juga merupakan tahap menghilangkan karakter atau angka yang nantinya akan mempengaruhi atau mengganggu dalam berjalannya pemrosesan teks. Berikut

ditampilkan hasil dari tahap *tokenizing* pada Tabel 4.

Tabel 4. Tokenize

Teks	Hasil
vaksin covid yang digunakan di indonesia sudah melalui uji klinis terbukti aman sudah melewati proses panjang	['vaksin', 'covid', 'yang', 'digunakan', 'di', 'indonesia', 'sudah', 'melalui', 'uji', 'klinis', 'terbukti', 'aman', 'sudah', 'melewati', 'proses', 'panjang']

3. *Stemming*

Stemming merupakan tahap proses pengubahan sebuah kata menjadi bentuk dasar yang tidak menggunakan imbuhan diawal maupun diakhir lalu sisipan dan juga perulangan dalam penggunaan kata tersebut. *Stemming* berjalan berdasarkan aturan yang dibuat, dengan memotong atau menghilangkan imbuhan sehingga tidak mustahil terdapat kata yang tidak konsisten dalam hasilnya, bentuk data yang tidak konsisten dari proses *stemming* dapat berupa *Overstemming* dimana pada kasus ini kata yang diolah terlalu banyak dipotong melebihi yang seharusnya sehingga arti katanya dapat hilang bila terjadi kasus ini. Berikut ditampilkan hasil dari tahap *stemming* pada Tabel 5.

Tabel 5. Stemming

Teks	Hasil
vaksin covid yang digunakan di indonesia sudah melalui uji klinis terbukti aman sudah melewati proses panjang	vaksin covid yang digunakan di indonesia sudah melalui uji klinis bukti aman sudah lewat proses panjang

4. *Filtering / Stopword Removal*

Tahap ini merupakan proses pemilihan kata-kata yang penting dalam NLP. *Filtering* memiliki dua metode berbeda yang dapat digunakan dalam melakukan pembersihan data, yang pertama adalah *Stoplist* dan yang kedua adalah *Wordlist*. Metode *Stoplist* merupakan metode menghilangkan kata yang tidak penting atau tidak bersifat deskriptif, dan *Wordlist* adalah metode yang berlawanan dengan *Stoplist* dimana dalam metode ini prosesnya adalah menyiapkan kata yang bermakna penting atau yang bersifat deskriptif, dimana hanya kata yang bersifat deskriptif saja yang akan disimpan untuk proses selanjutnya. Berikut ditampilkan hasil dari tahap *stopword removal* pada Tabel 6.

Tabel 6. Stopword Removal

Teks	Hasil
vaksin covid yang digunakan di indonesia sudah melalui uji klinis terbukti aman sudah melewati proses panjang	vaksin covid indonesia uji klinis aman lewat proses

Dataset yang sudah melewati tahapan praproses data akan berbentuk kumpulan kata yang memiliki bentuk dasar dan hanya berisi kata yang memiliki informasi penting. Berikut tampilan dataset saat dilakukan praproses data pada Gambar 2.

Text	remove_user
RT@Penyair_Berdiri : Vaksin COVID-19 yang digu...	RT : Vaksin COVID-19 yang digunakan di Indones...
remove_RT_angka_simbol	remove_tanda_baca
: Vaksin COVID- yang digunakan di Indonesia su...	Vaksin COVID yang digunakan di Indonesia suda...
remove_tanda_baca	tweet_tokenize_stemming_stopword
Vaksin COVID yang digunakan di Indonesia suda...	[vaksin, covid, indonesia, uji, klinis, bukti,...

Gambar 2. Praproses Dataset

C. Ekstraksi Data

Pada tahap ini ekstraksi data digunakan agar data komentar yang dimiliki dapat dibaca oleh mesin dengan mengubah data komentar berupa teks tersebut menjadi angka-angka sehingga memudahkan mesin untuk melakukan pengolahan data tersebut. Dalam tahap ini metode yang digunakan yaitu *Term Frequency Inverse Document Frequency* (TF-IDF). TF-idf adalah proses konversi sebuah teks menjadi angka, hal tersebut dilakukan karena *machine learning* dan arsitektur *deep learning* tidak mampu melakukan proses analisis pada input data berupa strings atau teks, sehingga membutuhkan angka sebagai input (Najiyah & Haryanti, 2021). Berikut ditampilkan pengolahan ekstraksi data menggunakan algoritma TF-IDF pada Gambar

```
df_tfidf.transpose().sample(1000)
```

	sentence 1	sentence 2	sentence 3	sentence 4	sentence 5	sentence 6	sentence 7
tapi	0.0	0.0	0.0	0.0	0.0	0.000000	0.0
remuk	0.0	0.0	0.0	0.0	0.0	0.000000	0.0
moodnya	0.0	0.0	0.0	0.0	0.0	0.000000	0.0
skill	0.0	0.0	0.0	0.0	0.0	0.000000	0.0
download	0.0	0.0	0.0	0.0	0.0	0.000000	0.0
gabisa	0.0	0.0	0.0	0.0	0.0	0.187857	0.0
jabatannya	0.0	0.0	0.0	0.0	0.0	0.000000	0.0
boostern	0.0	0.0	0.0	0.0	0.0	0.000000	0.0
gampang	0.0	0.0	0.0	0.0	0.0	0.000000	0.0
gencar	0.0	0.0	0.0	0.0	0.0	0.000000	0.0

1000 rows x 1300 columns

Gambar 3. Ekstraksi data TF-IDF

D. Pengujian

Sebelum digunakan untuk menganalisa sentimen, masing-masing algoritma dilakukan pengujian berdasarkan pembobotan nilai *accuracy*, *precision*, *recall* dan *f1-score*. Selain itu pengujian juga menggunakan metode *K-Fold Cross Validation* untuk memperkuat hasil pengujian, metode *K-Fold* yang dilakukan berupa pembelajaran data sebanyak 5 iterasi dan kemudian menghasilkan nilai terbaik dari *training accuracy* dan *test accuracy*.

E. Hasil Perbandingan

Setelah melakukan pembelajaran data pada dua algoritma *Convolutional Neural Network* dan *Naive Bayes* didapatkan hasil berupa kemampuan memprediksi analisa sentimen yang dinilai dari tingkat *accuracy*, *precision*, *recall* dan *f1-score* yang didapatkan berdasarkan *training* data komentar Twitter.

Accuracy didefinisikan sebagai tingkat jarak antara nilai prediksi algoritma dengan nilai sentimen berdasarkan labelnya. *Accuracy* digunakan untuk mengevaluasi banyaknya label prediksi yang sesuai dengan label aslinya. Semakin besar nilai akurasi, maka performa algoritma tersebut semakin baik. Berikut persamaannya:

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

Precision yaitu perbandingan jumlah komentar yang diprediksi benar dengan jumlah keseluruhan komentar yang telah di klasifikasi. *Precision* memiliki persamaan sebagai berikut :

$$\text{Precision} = \frac{TP}{TP + FN}$$

Recall, yaitu perbandingan jumlah komentar yang dikenali benar dengan jumlah seluruh komentar yang tepat dikenali. *Recall* memiliki rumusan sebagai berikut :

$$\text{Recall} = \frac{TP}{TP + FP}$$

F1-Score, yaitu rata-rata nilai yang didapatkan dari *Recall* dan *Precision* yang dibobotkan. *F1-Score* memiliki rumusan sebagai berikut :

$$F1 - \text{Score} = \frac{TP}{TP + \frac{1}{2}(FP + FN)}$$

Berdasarkan acuan-acuan tersebut, dilakukan perbandingan dari hasil keduanya untuk mengetahui algoritma apa yang lebih baik dalam mengolah data menjadi sebuah analisa sentimen. Algoritma yang lebih unggul dalam empat acuan tersebut akan digunakan modelnya untuk implementasi analisis sentimen komentar Twitter.

F. Analisa Sentimen

Pada tahap ini data dapat diolah untuk mencari informasi berupa sentimen dengan beberapa jenis polaritas yang terdapat dalam data komentar Twitter. Algoritma yang digunakan dalam analisis sentimen ini yaitu *Convolutional Neural Network* dan *Naïve Bayes*, keduanya akan diuji dalam hal kemampuan menganalisis sentimen.

1. *Convolutional Neural Network* 1D

Convolutional Neural Network 1D merupakan algoritma yang masuk ke dalam bagian *deep learning* yang cara kerjanya melakukan *training data* pada kumpulan data besar menggunakan parameter yang mengambil bentuk 1 dimensi sebagai *input*. Kemudian menggabungkan input tersebut dengan *filter* untuk mendapatkan *output* yang diinginkan.

Secara umum CNN terbagi menjadi 2 layer, layer ekstraksi dan layer klasifikasi. Layer ekstraksi terdapat dibagian awal arsitektur yang tersusun atas kumpulan layer dimana layer ini terdiri dari neuro-neuron yang terhubung dengan bagian sebelumnya. Sedangkan layer klasifikasi memiliki neuron yang terhubung dengan layer

lainnya tanpa ada bagian yang memisahkan antar neuron (Hermanto et al., 2021).

2. *Naïve Bayes Classifier*

Naïve Bayes merupakan teknik prediksi berbasis probabilistik sederhana yang berdasar pada penerapan teorema *Bayes* (atau aturan *Bayes*) dengan asumsi independensi (ketidaktergantungan) yang kuat (naïf). Dengan kata lain, *Naïve Bayes*, model yang digunakan adalah model fitur independen. *Naïve Bayes* memprediksi berdasarkan formula teorema *Bayes* sebagai berikut.

$$P[\text{Label} \setminus \text{Teks}] = \frac{P[\text{Teks} \setminus \text{Label}] * P[\text{Label}]}{P[\text{Teks}]}$$

Elemen utama *Naive Bayes Classifier* terdiri dari tiga aspek yaitu *priori*, *posteriori* dan *class condition probability* (Wibawa et al., 2019). Dalam teorema Bayes terdapat aturan dasar dimana hasil hipotesis pada Label diperkirakan atau dapat diprediksi berdasarkan keadaan teks setelah diamati .

3. Hasil dan Pembahasan

A. Pengujian algoritma

Untuk menentukan algoritma yang lebih baik dalam melakukan analisa sentimen maka dilakukan beberapa pengujian. Pengujian dilakukan pada algoritma *Convolutional Neural Network* dan *Naïve Bayes*. Dalam melakukan pengujian menggunakan jumlah pembagian data latih dan data *testing* yang berbeda untuk kemudian dilakukan pencarian nilai *precision*, *recall*, *f1-score*, dan *accuracy*. Pada algoritma CNN dibuat tiga pengujian dengan menggunakan tiga layer yang berbeda yaitu *Global Max Pooling Layer*, *Bidirectional Long Sort Term Memory layer*, dan *Long Sort Term Memory layer* untuk mengetahui nilai akurasi terbaik diantara tiga layer tersebut sehingga diketahui jenis layer apa yang memiliki nilai akurasi terbaik pada algoritma CNN. Tabel 7. dan Tabel 8. menunjukkan hasil dari pengujian yang dilakukan.

Kemudian pada Tabel 9. ditampilkan hasil pengujian nilai *best training accuracy* yang didapatkan menggunakan metode *K-Fold Cross Validation* pada algoritma CNN dan *Naïve Bayes*. Pengujian dilakukan dengan dua metode sehingga memperkuat analisa algoritma yang lebih unggul antara CNN dan *Naïve Bayes*.

Tabel 7. Pengujian Algoritma CNN *Global Max Pooling Layer* dan *Naïve Bayes*

Bobot Pengujian			CNN GlobalMaxPooling(%)				Naïve Bayes (%)			
Jumlah Data Latih (%)	Jumlah Data Uji (%)	Polaritas Sentimen	Precision	Recall	F1-Score	Accuracy	Precision	Recall	F1-Score	Accuracy
80	20	Pos	97	98	98	98	98	90	94	94
		Neg	98	97	98		91	98	94	
70	30	Pos	99	100	100	100	99	90	94	95
		Neg	100	99	100		91	99	95	
60	40	Pos	98	98	98	98	99	90	94	95
		Neg	98	98	98		91	99	95	

Tabel 8. Pengujian Algoritma CNN *BiLSTM Layer* dan *LSTM Layer*

Bobot Pengujian			CNN Bidirectional LSTM (%)				CNN LSTM (%)			
Jumlah Data Latih (%)	Jumlah Data Uji (%)	Polaritas Sentimen	Precision	Recall	F1-Score	Accuracy	Precision	Recall	F1-Score	Accuracy
80	20	Pos	98	96	97	97	99	95	97	97
		Neg	96	98	97		95	99	97	
70	30	Pos	97	97	97	97	98	95	97	97
		Neg	97	97	97		96	98	97	
60	40	Pos	96	96	96	96	97	96	96	96
		Neg	96	96	96		96	97	96	

Tabel 9. Pengujian *K-Fold* Algoritma CNN dan *Naïve Bayes*

K-Fold	CNN Global Max Pooling Layer		CNN BiLSTM Layer		CNN LSTM Layer		Naïve Bayes	
	Akurasi Train	Akurasi Test	Akurasi Train	Akurasi Test	Akurasi Train	Akurasi Test	Akurasi Train	Akurasi Test
K1	1	0.976923	1	0.984625	1	0.984615	0.978846	0.965385
K2	1	0.973077	1	0.973077	1	0.969231	0.981731	0.957692
K3	1	0.996154	1	0.969231	1	0.976923	0.981731	0.950000
K4	1	0.996154	1	0.984615	1	0.984615	0.976923	0.961538
K5	1	0.969231	1	0.980769	1	0.976923	0.980769	0.950000
Best Akurasi	1	0.996153	1	0.984615	1	0.984615	0.978846	0.965384
		84		38		38	15	15

B. Perbandingan Hasil Pengujian

Berdasarkan hasil pada pengujian menggunakan pembagian antara data latih dan data *testing* yang ditampilkan pada Tabel 7. dan Tabel 8. algoritma CNN dengan *Global Max Pooling layer* mendapatkan nilai rata-rata akurasi sebesar 98.66% dengan rata-rata nilai *precision*, *recall*, dan *f1-score* sebesar 98.33%, 98.33%, dan 98.66%, BiLSTM layer mempunyai nilai rata-rata akurasi sebesar 96.66% dan rata-rata nilai *precision*, *recall*, dan *f1-score* sebesar

96.66%, 96.66%, 96.66%, dan LSTM layer rata-rata akurasi sebesar 96.66% dengan rata-rata nilai *precision*, *recall*, dan *f1-score* sebesar 96.33%, 96.66%, 96.66%. Lalu algoritma *Naïve Bayes* mempunyai nilai rata-rata akurasi sebesar 94.66% dengan rata-rata nilai *precision*, *recall*, dan *f1-score* sebesar 94.83%, 94.33%, 94.33%.

Selain menggunakan metode pembagian data latih dan data *testing*, agar mendapatkan hasil yang lebih akurat dilakukan juga pengujian menggunakan metode *K-Fold*

Cross Validation untuk mengevaluasi model dengan efektif, secara sederhana metode merupakan metode memisahkan data latih dengan data *testing* secara bergantian sesuai dengan jumlah *K* atau jumlah iterasi yang diinginkan. Hasil pengujian dari metode *K-Fold* ditampilkan pada Tabel 9.

Setelah mendapatkan nilai rata-rata dari beberapa pembobotan pada Tabel 8. dan Tabel 9. maka didapatkan hasil bahwa algoritma CNN memiliki akurasi yang lebih baik dalam melakukan analisa sentimen dibandingkan dengan algoritma *Naïve Bayes*, pada pengujian algoritma CNN tersebut didapatkan hasil terbaik dengan nilai akurasi sebesar 100% pada model CNN yang menggunakan *Global Max Pooling layer* dengan besaran penggunaan data latih sebesar 70% dan data validasi sebesar 30%.

Pada pengujian berdasarkan besaran iterasi antara jumlah data latih dan data testing diketahui algoritma CNN mengungguli algoritma *Naïve Bayes* dalam menganalisa sentimen. Untuk memperkuat hasil pengujian tersebut maka digunakan metode *K-Fold Cross Validation*. *K-Fold Cross-validation* adalah metode statistik, dimana data dibagi menjadi dua subset, yaitu data pelatihan untuk proses pembelajaran dan pengujian data untuk validasi atau evaluasi, yang digunakan untuk mengevaluasi kinerja model, metode, atau algoritma. CV dapat dipilih berdasarkan ukuran himpunan data (Normawati & Ismi, 2019). Pada penerapan metode *K-Fold* terhadap algoritma yang diuji didapatkan hasil seperti yang ditampilkan pada Tabel 10. dimana algoritma CNN dengan *Global Max Pooling layer* kembali memiliki tingkat akurasi terbaik mengungguli dua *layer* CNN lainnya dengan nilai *best training accuracy* sebesar 100% dan *best testing accuracy* sebesar 99.61% sedangkan algoritma *Naïve Bayes* hanya mendapatkan nilai *best training accuracy* sebesar 97.88% dan *best testing accuracy* sebesar 96.53%

C. Analisa Sentimen

Karena algoritma CNN memiliki tingkat akurasi lebih baik dibandingkan algoritma *Naïve Bayes* dalam menganalisa sentimen, maka algoritma CNN akan digunakan untuk menganalisa sentimen masyarakat Indonesia mengenai kebijakan vaksin yang diterapkan pemerintah dengan menggunakan *Global Max Pooling layer* pada model training data. Proses analisa sentimen dilakukan dengan menggunakan *Python Notebook* pada *Google Collaboratory* yang akan digunakan untuk menjalankan model *training* data dengan

algoritma CNN dan kemudian model yang sudah di-*compile* akan digunakan untuk melakukan analisa sentimen secara otomatis pada file excel yang berisi komentar pengguna Twitter, file yang sebelumnya hanya berisi kolom komentar Twitter akan memiliki kolom tambahan berupa kolom sentimen setelah *script predict sentiment* dijalankan. Berikut adalah source code dari model algoritma CNN menggunakan *Global Max Pooling Layer*.

```
model_2 = tf.keras.Sequential([
    tf.keras.layers.Embedding(input_dim=5000, output_dim=256),
    tf.keras.layers.Conv1D(128, activation='relu', kernel_size=8),
    tf.keras.layers.GlobalMaxPooling1D(),
    tf.keras.layers.Dense(32, activation='relu'),
    tf.keras.layers.Dropout(0.5),
    tf.keras.layers.Dense(2, activation='softmax')
])
model_2.summary()

model_2.compile(loss='binary_crossentropy',
                optimizer=tf.optimizers.Adam(),
                metrics=['accuracy'])

# fit
history = model_2.fit(padded_train, Y_train, epochs=20,
                    verbose=2, callbacks=[early_stop], validation_data=(padded_test, Y_test))
```

Layer (type)	Output Shape	Param #
embedding_2 (Embedding)	(None, None, 256)	1280000
conv1d_2 (Conv1D)	(None, None, 128)	262272
global_max_pooling1d_2 (GlobalMaxPooling1D)	(None, 128)	0
dense_4 (Dense)	(None, 32)	4128
dropout_2 (Dropout)	(None, 32)	0
dense_5 (Dense)	(None, 2)	66

Total params: 1,546,466
 Trainable params: 1,546,466
 Non-trainable params: 0

Gambar 4. Source Code Model algoritma CNN

Pada model CNN terdiri dari beberapa lapisan yaitu lapisan *embedding*, konvolusi 1D, *Global Max Pooling*, termasuk beberapa lapisan *dense* dan *drop out*, dengan menggunakan fungsi optimasi *Adam*. *Adam Optimization* mampu meningkatkan akurasi dari kinerja jaringan syaraf tiruan (Jais et al., 2019).

Setelah model algoritma ter-*compile* proses selanjutnya adalah menjalankan *source code* untuk menganalisa sentimen dari dataset A yang berisi komentar Twitter yang belum dilabeli, *dataset* tersebut akan dilabeli secara otomatis berdasarkan pembobotan hasil *training data* yang dijalankan pada dataset B yang berisi komentar Twitter yang telah melalui tahap pra-proses data dan sudah diberikan label polaritas sentimen *positive* dan *negative*. Berikut *source code predict sentiment* yang akan menghasilkan label polaritas sentimen pada dataset A.


```

import tensorflow as tf
from tensorflow.keras.preprocessing.text import Tokenizer
from tensorflow.keras.preprocessing.sequence import pad_sequences

model_cnn_globalmaxpool = tf.keras.models.load_model('cnn_globalmaxpool.h5')

def convert_predict_cnn(predict):
    x = []
    for i in predict:
        if i == 0:
            x.append('negatif')
        elif i == 1:
            x.append('positif')
    return x

def preprocessing_cnn(X_train,X_test,Test=False):
    tokenizer = Tokenizer(num_words=5000, oov_token='')
    tokenizer.fit_on_texts(X_test)
    tokenizer.fit_on_texts(X_train)

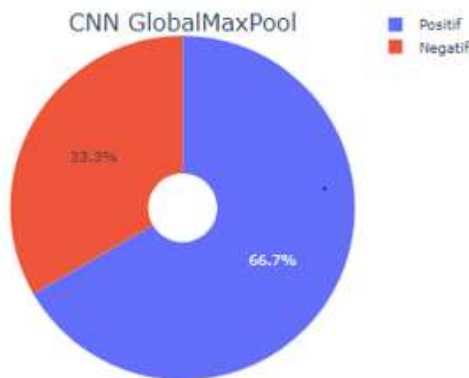
    sekuens_test = tokenizer.texts_to_sequences(X_test)
    padded_test = pad_sequences(sekuens_test)

    if Test == False:
        sekuens_train = tokenizer.texts_to_sequences(X_train)
        padded_train = pad_sequences(sekuens_train)
        return padded_train, padded_test
    else:
        return padded_test

padded_test = preprocessing_cnn(df_train,Tweet,df.Tweet,Test=True)
pred_cnn1 = model_cnn_globalmaxpool.predict(padded_test)
hasil_cnn1 = convert_predict_cnn(np.argmax(pred_cnn1, axis=1))
pred_cnn1 = model_cnn_globalmaxpool.predict(padded_test)
hasil_cnn1 = convert_predict_cnn(np.argmax(pred_cnn1, axis=1))
    
```

Gambar 5. Source Code Predict Sentiment

Hasil dari analisa sentimen berupa polaritas *positive* dan *negative* dan dapat dihitung persentase label masing-masing sentimen sebesar 66.7% untuk sentimen positif dan 33.3% untuk negatif, dari 1424 baris komentar diketahui terdapat 950 komentar berpolaritas positif dan 474 komentar berpolaritas negative. Hasil analisis berbentuk diagram dapat dilihat pada Gambar 6.



Gambar 6. Diagram Persentase polaritas sentiment

Untuk mengetahui kata yang dominan muncul pada masing-masing sentimen maka digunakan *wordcloud* untuk menampilkan beberapa kata yang terdapat dalam kolom sentimen, secara umum *wordcloud* bekerja dengan menampilkan kata dengan ukuran yang berbeda dimana semakin banyak jumlah kata muncul maka semakin besar ukuran kata dalam *wordcloud*. Berikut merupakan tampilan *wordcloud* dari kedua sentimen ditampilkan pada Gambar 7. dan 8..



Gambar 7. Wordcloud sentiment Positif



Gambar 8. Wordcloud sentiment Negatif

Dataset yang telah mendapatkan hasil analisis sentimen dapat disimpan menjadi file excel sebagai hasil akhir dari penggunaan algoritma CNN 1D dalam melakukan sentiment analisis. Gambar 9. memperlihatkan hasil final analisa sentimen pada komentar Twitter yang kemudian dapat disimpan kedalam format .XLSX.

df

51 to 75 of 1424 entries

index	Tweet	CNN GlobalMaxPool
50	ayo vaksin lebaran aman nyaman polisi indonesia polisirepublikindonesia polisiindonesia humaspolrespurbalingga	Positif
51	ayo vaksin lebaran aman polisi indonesia polisirepublikindonesia polisiindonesia humaspolrespurbalingga	Positif
52	ayo vaksin luur mari dukung sukses program juta vaksin kemenag laksana wilayah camat pakis minggu april wib lokasi yaspi pakis layan vaksin dosis covovac astrazeneca	Negatif
53	ayo vaksin mudik sehat keluarga covid	Positif
54	ayo vaksin mudik sukses juta vaksinasi booster pcnu kab bandung nu nahdlatululama nukabbandung pcnukabbandung sukseskanjutavaksinbooster kabbandung vaks ncovid ayovaks ncovid polrestabandung kemenagkabbandung dinkeskabbandung	Negatif
55	ayo vaksin sukses vaksinasi booster bukber disiplin prokes	Positif
56	ayo vaksin yg belum jgn tunda sakit jgn tunda kena razia ayo vaksin booster yg belum sesuai tipe vaksin anda sukses capa herd imunity	Positif
57	ayo vaksinasi tingkat daya tahan tubuh virus corona vaksin lupa terap protokol sehat m vaksin aman manfaat yuk vaksin dobsejahterakanpapua papuaindonesia	Positif
58	ayo warga kabupaten paser mari sukses program perintah juta vaksin booster jelang idul fitri h	Negatif
59	ayoo podo vaksin ben awak e sehat ora gampang kenek virus	Positif
60	ayoo vaksin indonesia sehat indonesia hebat	Positif
61	ayo vaksin bumi papua bebas covid kkbancamanbagipapua papuaindonesia	Positif
62	ayo vaksin ekonomi indonesia lonjak	Positif
63	ayo vaksin lengkapivaksininterapkanprokes indonesia negara	Positif

Gambar 9. Hasil Analisa Sentimen

4. Kesimpulan

Berdasarkan pengujian yang dilakukan, algoritma CNN dengan *Global Max Pooling layer* mendapatkan nilai rata-rata akurasi sebesar 98.66% dengan rata-rata nilai *precision*, *recall*, dan *f1-score* sebesar 98.33%, 98.33%, dan 98.66%, yang membuktikan keakuratannya dalam menganalisa sentimen lebih baik dari algoritma pembandingnya yaitu algoritma *Naïve Bayes* yang mempunyai nilai rata-rata akurasi sebesar 94.66% dengan rata-rata nilai *precision*, *recall*, dan *f1-score* sebesar 94.83%, 94.33%, 94.33%. Selain itu keakuratan algoritma CNN juga diperkuat dengan pengujian menggunakan metode *K-Fold Cross Validation* dimana algoritma CNN dengan *Global Max Pooling layer* kembali memiliki tingkat akurasi terbaik mengungguli dua layer CNN lainnya dengan nilai *best training accuracy* sebesar 100% dan *best testing accuracy* sebesar 99.61% sedangkan algoritma *Naïve Bayes* hanya mendapatkan nilai *best training accuracy* sebesar 97.88% dan *best testing accuracy* sebesar 96.53%.

Dari hasil analisa sentimen komentar seputar vaksinasi berbahasa Indonesia menggunakan algoritma CNN dengan *Global*

Max Pooling Layer dapat disimpulkan kebijakan vaksinasi ini mendapatkan respon yang cenderung positif dengan data komentar Twitter berjumlah 1424 baris, sebanyak 950 komentar berpolaritas positif dengan persentase 66.7% dan 33.3% sisanya sejumlah 474 komentar berpolaritas negatif. Selain itu berdasarkan data *Wordcloud* diketahui sebagian besar komentar bermuatan negatif berisi dengan kata-kata yang menyiratkan efek samping dari vaksinasi terutama jenis vaksinasi *booster*.

Referensi

- Ayumi, V., & Nurhaida, I. (2021). Klasifikasi Chest X-Ray Images Berdasarkan Kriteria Gejala Covid-19 Menggunakan Convolutional Neural Network. *JSAI: Journal Scientific and Applied Informatics*, 4(2), 147–153.
- Dhika, H., Kurnianda, N. R., Irfansyah, P., & Ananta, W. (2020). Model Prediksi Jenis Hewan dengan Metode Convolution Neural Network. *Jurnal Ilmiah Teknik Informatika*, 9(1), 31–40.
- Dwianto, E., Sadikin, M., Informatika, J. T., Komputer, F. I., & Buana, U. M. (2021). Analisis Sentimen Transportasi Online

- pada Twitter Menggunakan Metode Klasifikasi Naïve Bayes dan Support Vector Machine. *Jurnal Ilmiah Teknik Informatika*, 10(1), 94–100.
- Hermanto, D. T., Setyanto, A., & Luthfi, E. T. (2021). Algoritma LSTM-CNN untuk Binary Klasifikasi dengan Word2vec pada Media Online. *Creative Information Technology Journal*, 8(1), 64–77. <https://doi.org/10.24076/citec.2021v8i1.264>
- Ihsan, C. N. (2021). DoubleClick : Journal of Computer and Information Technology Klasifikasi Data Radar(Ihsan) | 115 DoubleClick : Journal of Computer and Information Technology E-ISSN : 2579-5317 116 | Klasifikasi Data Radar(Ihsan). *Journal of Computer and Information Technology*, 4(2), 115–121.
- Jais, I. K. M., Ismail, A. R., & Nisa, S. Q. (2019). Adam Optimization Algorithm for Wide and Deep Neural Network. *Knowledge Engineering and Data Science*, 2(1), 41. <https://doi.org/10.17977/um018v2i12019p41-46>
- Khan, M., Khan, H., Khan, S., & Nawaz, M. (2020). Epidemiological and clinical characteristics of coronavirus disease (COVID-19) cases at a screening clinic during the early outbreak period: a single-centre study. *Journal of Medical Microbiology*, 69(8), 1114–1123. <https://doi.org/10.1099/jmm.0.001231>
- Najiyah, I., & Haryanti, I. (2021). SENTIMEN ANALISIS COVID-19 DENGAN METODE. *Jurnal Responsif : Riset Sains Dan Informatika*, 3(1), 100–111. <https://doi.org/https://doi.org/10.51977/jti>
- Nasichuddin, M. A., Adji, T. B., & Widyawan, W. (2018). Performance Improvement Using CNN for Sentiment Analysis. *IJITEE (International Journal of Information Technology and Electrical Engineering)*, 2(1), 9–14. <https://doi.org/10.22146/ijitee.36642>
- Normawati, D., & Ismi, D. P. (2019). K-Fold Cross Validation for Selection of Cardiovascular Disease Diagnosis Features by Applying Rule-Based Datamining. *Signal and Image Processing Letters*, 1(2), 23–35. <https://doi.org/10.31763/simple.v1i2.3>
- Qudsi, D. H., Lubis, J. H., Syaliman, K. U., & Najwa, N. F. (2021). Analisis Sentimen Pada Data Saran Mahasiswa Terhadap Kinerja Departemen Di Perguruan Tinggi Menggunakan Sentiment Analysis In The Student's Reviews Of College Department Performance Using. *Jurnal Teknologi Informasi Dan Ilmu Komputer (JTIK)*, 8(5), 1067–1076. <https://doi.org/10.25126/jtiik.202184842>
- Rhanoui, M., Mikram, M., Yousfi, S., & Barzali, S. (2019). A CNN-BiLSTM Model for Document-Level Sentiment Analysis. *Machine Learning and Knowledge Extraction*, 1(1), 832–847. <https://doi.org/10.3390/make1030048>
- Santoso, A., & Ariyanto, G. (2018). Implementasi Deep Learning berbasis Keras untuk Pengenalan Wajah. *Emitor: Jurnal Teknik Elektro*, 18(1), 15–21. <https://doi.org/10.23917/emitor.v18i01.6235>
- Susilo, A., Rumende, C. M., Pitoyo, C. W., Santoso, W. D., Yulianti, M., Herikurniawan, H., Sinto, R., Singh, G., Nainggolan, L., Nelwan, E. J., Chen, L. K., Widhani, A., Wijaya, E., Wicaksana, B., Maksum, M., Annisa, F., Jasirwan, C. O. M., & Yuniastuti, E. (2020). Coronavirus Disease 2019: Tinjauan Literatur Terkini. *Jurnal Penyakit Dalam Indonesia*, 7(1), 45. <https://doi.org/10.7454/jpdi.v7i1.415>
- Wibawa, A. P., Kurniawan, A. C., Murti, D. M. P., Adiperkasa, R. P., Putra, S. M., Kurniawan, S. A., & Nugraha, Y. R. (2019). Naïve Bayes Classifier for Journal Quartile Classification. *International Journal of Recent Contributions from Engineering, Science & IT (IJES)*, 7(2), 91. <https://doi.org/10.3991/ijes.v7i2.10659>