

SURAT PERNYATAAN

Orisinalitas Karya Lomba Penambangan Data (Data Mining)
GEMASTIK XIV 2021

Nama Tim : Yaudahlah

Judul Karya : Pendekatan Deep Learning untuk Mengklasifikasikan Opini Publik mengenai Vaksinasi COVID-19 di Indonesia

Anggota Tim:

1. Kaenova Mahendra Auditama (1301190324)
2. Ananda Affan Fattahila (1301194175)
3. Fendi Irfan Amorokhman (1301191447)

Kami dengan identitas di atas menyatakan bahwa karya kami yang berjudul Pendekatan Deep Learning untuk Mengklasifikasikan Opini Publik mengenai Vaksinasi COVID-19 di Indonesia dan diikutsertakan dalam Lomba Penambangan Data (Data Mining) GEMASTIK XIV 2021 adalah karya asli, tidak menjiplak/mencontek karya orang lain dan melanggar hak cipta. Karya tersebut belum pernah dinyatakan sebagai finalis lomba penambangan data sebelumnya dan karya tersebut belum pernah diikutsertakan dalam perlombaan apapun. Pada dokumen ini kami lampirkan pemeriksa plagiarisme menggunakan Turnitin.

Demikian surat pernyataan ini kami buat dengan sebenar-benarnya. Apabila dikemudian hari kami terbukti melanggar ketentuan tersebut, maka kami bersedia didiskualifikasi, baik sebagai peserta maupun pemenang.

Menyetujui,
Dosen Pembimbing



(Siti Sa'adah, S.T., M.T.)
NIDN. 0411028604

Ketua Tim Yaudahlah



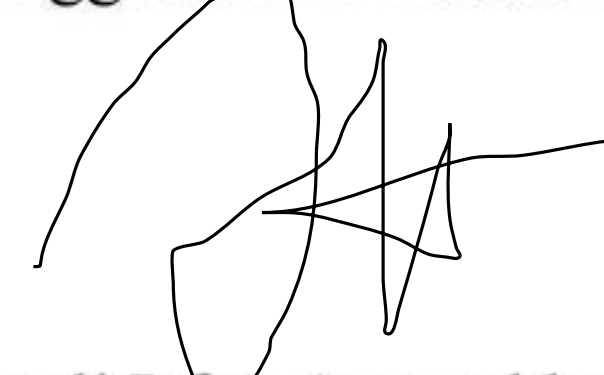
(Kaenova Mahendra Auditama)
NIM. 1301190324

Anggota Tim Yaudahlah



(Ananda Affan Fattahila)
NIM. 1301194175

Anggota Tim Yaudahlah



(Fendi Irfan Amorokhman)
NIM. 1301191447

Gemastik XIV: Yaudahlah

by kaenova 1

Submission date: 28-Jul-2021 01:06PM (UTC-0400)

Submission ID: 1625041749

File name: Proposal_Gemastik_6.pdf (1.09M)

Word count: 4098

Character count: 25300



PENDEKATAN DEEP LEARNING UNTUK MENGKLASIFIKASIKAN OPINI PUBLIK MENGENAI VAKSINASI COVID-19 DI INDONESIA

Disusun oleh:

Tim Yaudahlah

Anggota:

Ananda Affan Fattahila

Fendi Irfan Amorokhman

Kaenova Mahendra Auditama

Asal Universitas:

Telkom University



ABSTRAK

Virus COVID-19 dinyatakan sebagai pandemi pada bulan Maret 2020 dan pada bulan Juli 2021 varian COVID-19 Delta memasuki Indonesia yang memiliki probabilitas penularan yang tinggi. Kesehatan menjadi salah satu faktor yang krusial karena harus melakukan vaksinasi untuk menciptakan kekebalan tubuh pada masyarakat Indonesia. Masyarakat menjadi gelisah, panik dan emosi, sehingga berusaha mengeluarkan opini misalnya melalui unggahan text pada Twitter. Hal ini dapat dilakukan untuk mengambil analisis sentimen dengan menggunakan metode *deep learning* seperti BERT, CNN-LSTM, BERT-LSTM. Dengan menggunakan tiga model dapat menghasilkan prediksi yang lebih tervalidasi untuk menentukan hasil analisis sentimen. Dalam penelitian ini menghasilkan nilai sentimen yang cenderung negatif pada beberapa pola pikir opini publik. Salah satu bentuk opini publik mengenai keefektifan vaksin menghasilkan persentase BERT 70% , CNN-LSTM 58%, BERT-LSTM 67% dari total 8411 dataset. Dapat dilihat dari hasil prediksi ketiga model menyatakan bahwa persentase sentimen negatif lebih besar daripada sentimen positif. Sentimen negatif dapat menjadi sebuah bahan pembelajaran untuk memperbaiki Indonesia agar lebih maju dalam mengakhiri pandemi COVID-19.

Kata Kunci: Pemrosesan Bahasa, Twitter, Opini, COVID-19, BERT, CNN-LSTM, BERT-LSTM

1. PENDAHULUAN

1.1. Latar Belakang

Pada 11 Maret 2020, WHO atau yang dikenal sebagai organisasi kesehatan dunia menyatakan bahwa virus corona atau COVID-19 sebagai pandemi dunia. Pengambilan keputusan ini berdasarkan kasus positif di luar China yang meningkat tiga belas kali lipat di 114 negara [1]. *Coronavirus Disease 2019* atau yang kita kenal dengan sebutan COVID-19 yang disebabkan oleh virus corona memiliki kemampuan reproduksi lebih tinggi dari virus SARS [2]. Sampai saat ini kita tahu bahwa pandemi ini belum berakhir, bahkan minggu lalu (21/7/2021) menandai minggu keempat berturut-turut peningkatan kasus COVID-19 secara global, dengan peningkatan tercatat di semua kecuali satu wilayah dari enam wilayah WHO [3]. Virus corona COVID-19 sudah mengalami beberapa kali proses mutasi yang membuat virus ini memiliki beberapa variasi. Virus varian delta saat ini tengah ramai karena membuat lonjakan yang tinggi terhadap angka positif dan kematian. Delta sekarang sudah ada di lebih 104 negara termasuk Indonesia [3].

Saat masa pandemi ini COVID-19, sektor kesehatan merupakan sektor yang sangat krusial/penting [20]. Hal ini sudah tidak dapat dipungkiri karena COVID-19 merupakan masalah besar di bidang kesehatan yang dapat menyebabkan kematian. Sehingga penanggulangan COVID-19 pemerintah membuat kebijakan baru seperti Pembatasan sosial berskala Besar, Pemberlakuan Pembatasan Kegiatan Masyarakat, Protokol Kesehatan, dan masih banyak kebijakan lainnya [21]. Namun dalam hal mencegah penyebaran virus COVID-19 alat yang paling ampuh untuk melakukan hal tersebut adalah Vaksin. Tujuan dari Vaksin adalah untuk memberikan kekebalan terhadap penyakit [4].

Untuk menjelaskan dan memprediksi terkait respons emosional publik, khususnya sentimen vaksinasi COVID-19 di Indonesia, terdapat 40.628 postingan publik yang dapat dianalisis dari tanggal 19 Januari 2021 hingga 19 Juli 2021 di Twitter, salah satu media sosial populer di Indonesia dengan pengguna aktif 19,5 Juta. Pada penelitian ini diambil postingan publik yang membahas COVID-19 mulai dari segi keefektifan, harga dan merk yang digunakan. Model yang bangun untuk melakukan klasifikasi pada penelitian ini Bidirectional Encoder Representations from Transformers (BERT), CNN-LSTM dan BERT-LSTM yang menghasilkan klasifikasi prediksi terhadap opini - opini publik terkait vaksin COVID-19 di Indonesia.

1.2. Tujuan dan Manfaat

Dalam penelitian ini memiliki beberapa tujuan dan manfaat sebagai berikut:

1. Mendeskripsikan metode dengan akurasi terbaik dalam sentimen vaksin COVID-19 di Indonesia.
2. Mendeskripsikan opini masyarakat Indonesia terhadap keefektifan vaksin COVID-19.
3. Mendeskripsikan opini masyarakat Indonesia terhadap pemberian vaksin COVID-19 gratis dan berbayar di Indonesia. [36]
4. Mendeskripsikan opini masyarakat Indonesia terkait jenis vaksin COVID-19 di Indonesia.

[24]

1.3. Batasan Masalah

Dalam penelitian ini terdapat beberapa hal yang merupakan batasan masalah. Hal ini dilakukan guna memberikan gambaran terkait ruang lingkup penelitian yang dilakukan:

1. Menggunakan text yang didapat dari Twitter dengan filter bahasa Indonesia.
2. Menggunakan text yang didapat dari Twitter dengan filter mematikan retweet.
3. Maksimal token pada model yang diproses adalah 128 token.
4. Untuk data Twitter, data yang diambil dimulai dari 19 Januari 2021 - 19 Juli 2021.

2. METODOLOGI PENELITIAN

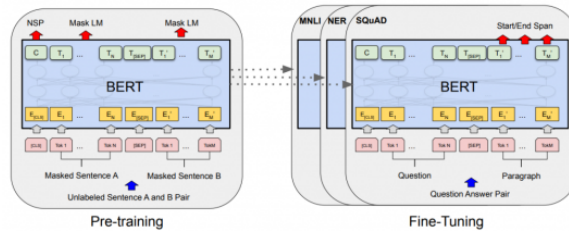
2.1. Analisis Sentimen

Analisis Sentimen adalah teknik yang sangat penting untuk menggali emosi publik dan merupakan area penelitian multidisiplin dalam perkembangan interaksi media sosial dan menangani volume data yang besar [5]. Sentimen publik merupakan data yang sangat krusial dalam sebuah penelitian untuk menganalisis sudut pandang

emosi, pemikiran, dan opini seseorang terhadap suatu masalah. Virus corona merupakan sebuah pandemi global, sehingga banyak peneliti yang tengah mencari tahu mengenai virus ini, mulai dari jenis, vaksin, dampak dan masih banyak hal lain. Selama masa pandemi banyak masyarakat Indonesia yang mengekspresikan emosi, pikiran serta pendapat mereka terkait COVID-19 melalui twitter. Saat ini, twitter juga menjadi salah satu media sosial yang kuat untuk berbagi informasi terutama informasi publik melalui metode online [5]. Motif utama dari penelitian ini adalah untuk mengidentifikasi tweet positif, negatif, dan memvisualisasikan masalah untuk meningkatkan rencana pembuat kebijakan untuk memerangi pandemi ini [6].

2.2. Bidirectional Encoder Representations from Transformers (BERT)

Bidirectional Encoder Representations from Transformers (BERT), teknik berbasis jaringan saraf untuk pra-pelatihan pemrosesan bahasa alami, telah banyak diterapkan dalam analisis sentimen [7]. Model BERT dapat disesuaikan dengan lapisan input dan output yang tepat untuk membuat model *state-of-the-art* dalam berbagai tugas analisis teks [8]. BERT merupakan sebuah model representasi bahasa baru yang dapat menghasilkan model *pre-train* representasi *bidirectional* dari suatu teks yang memiliki kelas dengan bersamaan mengkondisikan dari kedua subjek pada semua layer. Hal ini yang menyebabkan model BERT yang sudah di train, bisa disesuaikan misalnya dengan menambah satu layer output saja.



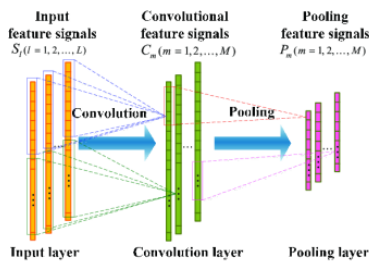
Gambar 2.1 Ilustrasi BERT Pretrain dan Fine-Tuning [8]

Pada model BERT terdapat dua langkah yang penting. Langkah itu adalah *pre-train* dan *fine-tuning*. Pada langkah *Pre-train* akan dilakukan *training* pada model terhadap data yang tidak memiliki kelas. Langkah *fine-tuning* diawali dengan parameter pada *pre-train* dan semua parameternya disesuaikan dengan menggunakan data yang sudah berlabel.

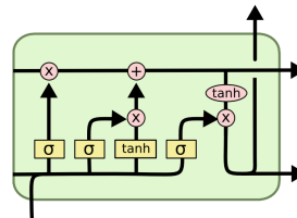
2.3. Convolutional Neural Network (CNN) - Long Short Term Memory (LSTM)

Convolutional Neural Network (CNN) merupakan salah satu tipe *neural network* yang sangat baik jika dihadapkan oleh data spasial. Layer *Convolutional* hanya digunakan untuk koneksi sekarang dari layer sebelumnya, lokal *neuron* terhubung ke *neurons* di layer selanjutnya. Metode ini berguna untuk meningkatkan pengetahuan pada layer untuk memahami lebih general terhadap input yang masuk.

Long Short Term Memory (LSTM) merupakan salah satu pengembangan *neural network* yang dapat digunakan untuk pemodelan data *time series* [9]. LSTM juga dapat diartikan sebagai jaringan saraf dengan arsitektur yang mudah beradaptasi, sehingga bentuknya dapat disesuaikan, tergantung pada aplikasi yang di buat [10].



Gambar 2.2 Convolutional Neural Network



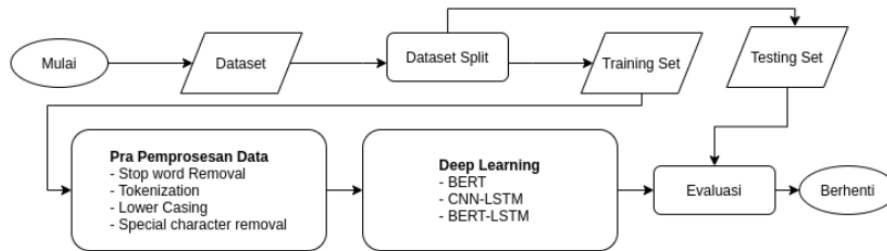
Gambar 2.3 Model LSTM

Pada LSTM terdapat beberapa steps, yaitu *define network*, *compile network*, *fit network*, *evaluate network*, dan *make prediction*. Dengan menggunakan LSTM yang dapat melakukan penentuan terhadap suatu nilai yang dimana akan dijadikan sebagai keluaran yang relevan terhadap masukan yang diberikan.

Pada CNN input berupa vektor-vektor yang mana merepresentasikan setiap kata lalu akan diambil. Lalu akan dipilih kernel pada vektor tersebut dengan ukuran tertentu yang disebut *kernel size* selanjutnya akan dilakukan *pooling* terhadap vektor hasil dari layer *convolutional*.

2.4. Desain Sistem

Pada penelitian ini menggunakan pendekatan *deep learning* untuk menangani data yang kompleks dan untuk mendapatkan akurasi yang baik. Alur kerja yang digunakan meliputi tahap pengumpulan dataset, *dataset split*: *training set* dan *testing set*, pra pemrosesan data, *deep learning* dan evaluasi, dengan gambaran sistem sebagai berikut :



Gambar 2.4 Sistem yang Dibangun

2.5. Dataset

Pada pengumpulan dataset dilakukan penggabungan dari beberapa sumber data untuk melatih mesin agar dapat memberikan prediksi yang baik terhadap sentimen masyarakat di twitter. Data yang pertama digunakan adalah data yang telah dikembangkan oleh IndoLEM <https://github.com/indolem/indolem/tree/main/sentiment/data> [11]. Data yang kedua merupakan data yang dikembangkan oleh indo SMSA https://github.com/indobenchmark/indonlu/tree/master/dataset/smsa_doc-sentiment-prosa [18].

2.6. Pra Pemrosesan Data

Pra-pemrosesan data adalah proses pembersihan dan penghapusan zat non-tekstual [12]. Beberapa teknik pra pemrosesan data yang diterapkan sebagai berikut:

2.6.1. Stop Word Removal

Stopword Removal adalah proses pembuangan kata yang tidak memiliki arti atau tidak relevan [13]. Adapun list stopwords yang digunakan pada preprocessing di penelitian ini berasal dari <https://github.com/datascienceid/stopwords-bahasa-indonesia>.

Kalimat	Hasil Stop Word Removal Kalimat
'Aku sangat menantikan vaksinasi di Indonesia'	'menantikan vaksinasi Indonesia'

2.6.2. Tokenization

Tokenizing merupakan tahap pemotongan pada setiap kata dalam sebuah kalimat atau parsing dengan menggunakan spasi sebagai delimiter dan akan menghasilkan token berupa kata [14].

Kalimat	Hasil Tokenisasi Kalimat
'Vaksin sangat dibutuhkan untuk tercapainya imunitas'	[2, 11087, 310, 2685, 90, 15380, 24272, 3]

2.6.3. Lower Casing

Teknik ini digunakan untuk melakukan *lower case* pada semua karakter yang ada pada teks.

Kalimat	Hasil Lower Casing Kalimat
'Saya seseorang yang sudah divaksin. Saya tidak sakit.'	'saya seseorang yang sudah divaksin. saya tidak sakit.'

2.6.4. Special Character Removal

Teknik ini digunakan untuk menghapus Character termasuk tanda baca, emoticon dll pada teks.

Kalimat	Hasil Special Character Removal Kalimat
'Saya tidak suka kamu -_- 🤔'	'Saya tidak suka kamu'

2.6.5. Penggantian kata *typo* dan *slang*

Teknik ini digunakan untuk mengganti kata *slang* dan *typo* dari kalimat menjadi kalimat yang baku dan lebih umum. Pada penelitian ini digunakan kamus dari sumber <https://github.com/nasalsabila/kamus-alay> dan <https://github.com/fendiirfan/Kamus-Alay>.

Kalimat	Hasil Penggantian Typo dan Slang Kalimat
'gue udah di vaksin, loe kapan?'	'aku udah di vaksin, kamu kapan?'

2.7. Arsitektur Model

Pada penelitian ini digunakan tiga model yang memiliki arsitektur model masing-masing, hal ini bertujuan untuk mencari metode dengan akurasi terbaik. Adapun arsitektur dari setiap model yang digunakan sebagai berikut:

2.7.1. BERT



Gambar 2.5 Arsitektur Pemodelan BERT

Pada model ini, digunakan IndoBERT sebagai BERT *base* dalam arsitektur model ini. IndoBERT adalah sebuah model BERT yang sudah pernah dilatih dengan menggunakan Masked Language Modelling [17]. Model IndoBERT tersebut dilakukan penyesuaian terhadap keluaran yang diinginkan, pada hal ini merupakan sebuah klasifikasi dengan 2 kategori, sehingga dibutuhkan satu *layer* tambahan *Dense* dengan aktivasi *Softmax*.

2.7.2. CNN-LSTM



Gambar 2.6 Arsitektur Pemodelan CNN-LSTM

Pada model ini, masukan ke model merupakan kalimat yang selanjutnya akan melalui *tokenizer layer* yang bertujuan untuk membagi kalimat menjadi token token yang memiliki output dimension 128 lalu melewati *embedding layer* dari *library* keras. *Layer* selanjutnya adalah *layer* convolutional bertipe *one-dimensional* dengan *activation function* *relu*. Selanjutnya akan melewati LSTM Layer sebesar 128 layer lalu melewati *dense layer* sebanyak 128 lalu dilakukan *dropout* sebanyak 0.5 lalu di-*outputkan* dengan *dense layer* dengan *activation function* *sigmoid*.

2.7.3. BERT-LSTM



Gambar 2.7 Arsitektur Pemodelan BERT-LSTM

Pada model ini, digunakan IndoBERT yang ditambahkan LSTM sebesar 128 *layer*. Sebelum disesuaikan untuk mengklasifikasikan, diberikan *layer* untuk melakukan normalisasi pada keluaran LSTM dan setelah itu dilakukan pengklasifikasian pada *layer Dense* dengan aktivasi *Softmax*.

2.8. Evaluasi

Dalam proses training dilakukan penerapan aturan parameter yang berpengaruh terhadap hasil evaluasi, dimana aturan parameter yang digunakan pada setiap model deep learning yang digunakan di buat sama, dengan parameter dibawah ini :

Tabel 2.1 Parameter Model

Parameter	Nilai
Max Sequence Length	128
Loss function	Categorical Cross-entropy
Optimizer	Adam
Epoch	5
Training / Validation Split	0.8
Learning Rate	3e-6 (BERT, BERT-LSTM) dan 1e-4 (CNN-LSTM)

Dengan menggunakan dataset, arsitektur model dan parameter yang digunakan. Dapat dilakukan implementasi kode dengan *Environment Google Collab* dan menggunakan GPU. Proses training memakan waktu dan hasil sebagai berikut :

Tabel 2.2 Performansi Model terhadap Data Training

Model	Akurasi	Loss	Precision	Recall	F1
CNN-LSTM	0.88	0.31	0.88	0.88	0.88
BERT	0.97	0.06	0.97	0.97	0.98
BERT-LSTM	0.96	0.11	0.96	0.96	0.96

Tabel 2.3 Performansi Model terhadap Data Validasi

Model	Akurasi	Loss	Precision	Recall	F1
CNN-LSTM	0.83	0.38	0.83	0.83	0.83
BERT	0.80	0.70	0.80	0.80	0.82
BERT-LSTM	0.79	0.62	0.79	0.79	0.79

Tabel 2.4 Performansi Model terhadap Data Testing

Model	Akurasi	Loss	Precision	Recall	F1
CNN-LSTM	0.70	0.73	0.70	0.70	0.81
BERT	0.80	0.73	0.78	0.78	0.80
BERT-LSTM	0.80	0.69	0.78	0.78	0.80

Hasil weight dan model yang sudah di training dapat diunduh pada link berikut: bit.ly/YaudahlahModels

Pembuatan sistem serta pembuatan model dapat dilihat pada link berikut: bit.ly/YaudahlahNotebooks

Data yang digunakan dapat dilihat pada link berikut: bit.ly/YaudahlahData

3. DATA DAN ANALISIS

Analisis dilakukan dengan melakukan perhitungan semantik Masyarakat Indonesia terhadap pemberian vaksin oleh pemerintah yang telah dilaksanakan atau masih direncanakan untuk di implementasi. Pada implementasi pengambilan data yang digunakan, sebagai bahan analisis terhadap sentimen masyarakat terkait opini publik terhadap vaksinasi COVID-19 di Indonesia. Data yang dikumpulkan berasal dari platform Twitter dengan menggunakan twint (<https://github.com/twintproject/twint>). Data yang ambil berjumlah 40.628 dengan filter mengecualikan retweet sehingga data yang diperoleh merupakan opini publik itu sendiri dan filter bahasa Indonesia. Dari data tersebut dapat dilakukan prediksi dengan model yang sudah di training untuk menentukan klasifikasi hasil sentimen publik yang hasilnya akan digambarkan menjadi diagram berikut:

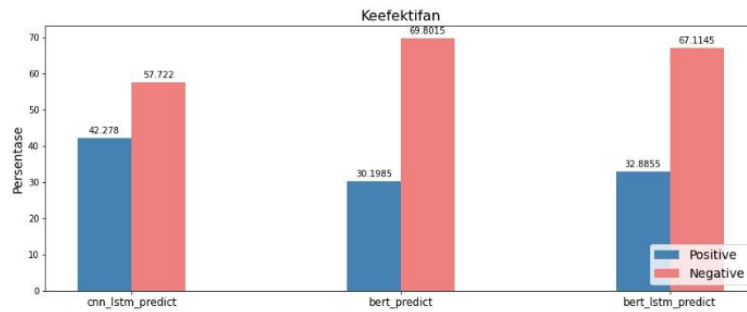
3.1. Opini Masyarakat Indonesia Terhadap Keefektifan Vaksin COVID-19

Keefektifan merupakan suatu keadaan yang berpengaruh, kemandirian atau suatu keberhasilan [14]. Sedangkan keefektifan vaksin dapat diartikan menjadi kemandirian atau keberhasilan dari vaksin COVID-19. Kepercayaan terkait keefektifan vaksin sangatlah penting karena Vaksin merupakan menjadi salah satu bentuk pencegahan terhadap virus COVID-19. Di Indonesia pemberian vaksin sendiri tergolong masih belum lama, dimana vaksin mulai diberikan pada awal tahun 2021. Sehingga banyak masyarakat yang memberikan opini terkait Keefektifan Vaksin. Opini masyarakat yang digunakan untuk melakukan analisis sentimen diambil dari twitter dengan keyword "vaksin", "vaksinasi", "vaksinisasi", "covid", "corona", "covid-19", "covid19", "berguna", "bermanfaat", "berhasil", "menghidupkan", "gagal", "mematikan", "tidak berguna" dan "merugikan" dengan contoh hasil dataset sebagai berikut :

Tabel 3.1 Sampel Hasil Prediksi Model terhadap Keefektifan Vaksin COVID-19

Kalimat	Prediksi BERT	Prediksi CNN - LSTM	Prediksi BERT-LSTM
bantu pemerintah melawan covid cara patuhi protokol kesehatan	1	1	1
seneng banget berhasil membujuk bapak vaksin	1	1	1
orang udah vaksin kali bisa kena covid yaa bingung banget anj tetangga gua kena covid udah vaksin kalijadi vaksin berguna ga sih	-1	-1	-1

Dalam dataset tersebut terlihat bahwa kalimat-kalimat yang digunakan merupakan kalimat yang didapatkan dari hasil crawling twitter dan belum memiliki label kategori positif atau negatif. Setelah dilakukan prediksi kategori, dapat dilakukan pengelompokan sesuai hasil kategorinya dan didapatkan persentase sebagai berikut :



Gambar 3.1 Persentase Sentimen terhadap Keefektifan Vaksin

Dari gambar diatas terlihat bahwa masyarakat cenderung memiliki sentimen negatif terhadap keefektifan vaksin COVID-19 di Indonesia. Hal tersebut, dapat terlihat dari persentase yang dihasilkan oleh model CNN-LSTM 58% negatif dan 42% positif, sedangkan model BERT 70% negatif dan 30% positif, dan BERT-LSTM 67% negatif dan 33% positif dari total 8411 dataset. Hal ini menunjukkan bahwa model CNN-LSTM, BERT, dan BERT-LSTM dengan menggunakan parameter pada tabel 2.1 menunjukan hasil yang serupa kasus ini.

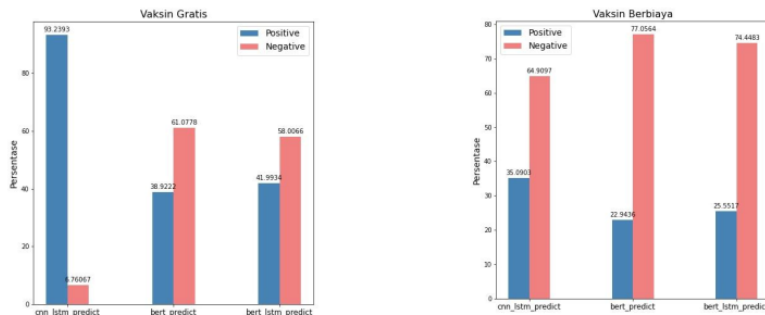
3.2. Opini Masyarakat Indonesia Terhadap Pemberian Vaksin Gratis dan Berbayar

Vaksin COVID-19 gratis dan berbayar menjadi sebuah hal yang lumayan krusial dimana perbedaan akan menyebabkan kesenjangan sosial dan kelas sosial baru. Namun hal ini juga memiliki sisi lain yang mungkin akan berdampak terhadap cepatnya proses vaksinasi. Sehingga hal ini mengundang opini masyarakat. Opini masyarakat yang digunakan untuk analisis sentimen pada opini masyarakat indonesia terhadap pemberian vaksin COVID-19 gratis dan berbayar menggunakan keyword: "vaksin", "vaksinasi", "vaksinisasi", "covid", "corona", "covid-19", "covid19", "gratis", "murah", "mahal", "bayar", "berbayar", "tanpa biaya" dengan contoh hasil dataset sebagai berikut :

Tabel 3.2 Sampel Hasil Prediksi Model terhadap Vaksin Gratis dan Berbayar COVID-19.

Kalimat	Prediksi BERT	Prediksi CNN - LSTM	Prediksi BERT-LSTM
ayo cari vaksin mumpung gratis	1	1	1
udh vaksin gratis dapet sembako mantap emang	1	1	1
ih obat covid tu jatohnya mahal	-1	-1	-1

Setelah dilakukan prediksi kategori, dapat dilakukan pengelompokan sesuai hasil kategorinya dan didapatkan persentase sebagai berikut :



Gambar 3.2 Persentase Sentimen terhadap Pemberian Vaksin Secara Gratis dan berbayar

Dari gambar diatas terlihat bahwa masyarakat cenderung memiliki sentimen negatif terhadap vaksin berbayar. Hal tersebut, dapat terlihat dari persentase yang dihasilkan oleh model CNN-LSTM 65% negatif dan 35% positif, sedangkan model BERT 77% negatif dan 23% positif, dan BERT-LSTM 74% negatif dan 26% positif dari total 5177 dataset. Sedangkan terhadap vaksin gratis cenderung negatif untuk BERT dan BERT-LSTM dan positif untuk CNN-LSTM Hal tersebut, dapat terlihat dari persentase yang dihasilkan oleh model CNN-LSTM 7% negatif dan 93% positif, sedangkan model BERT 61% negatif dan 39% positif, dan BERT-LSTM 58% negatif dan 42%

positif dengan total 5483 dataset. Hal ini menunjukkan bahwa model CNN-LSTM, BERT, dan BERT-LSTM dengan menggunakan parameter pada tabel 2.1 menunjukkan hasil yang berbeda pada kasus ini.

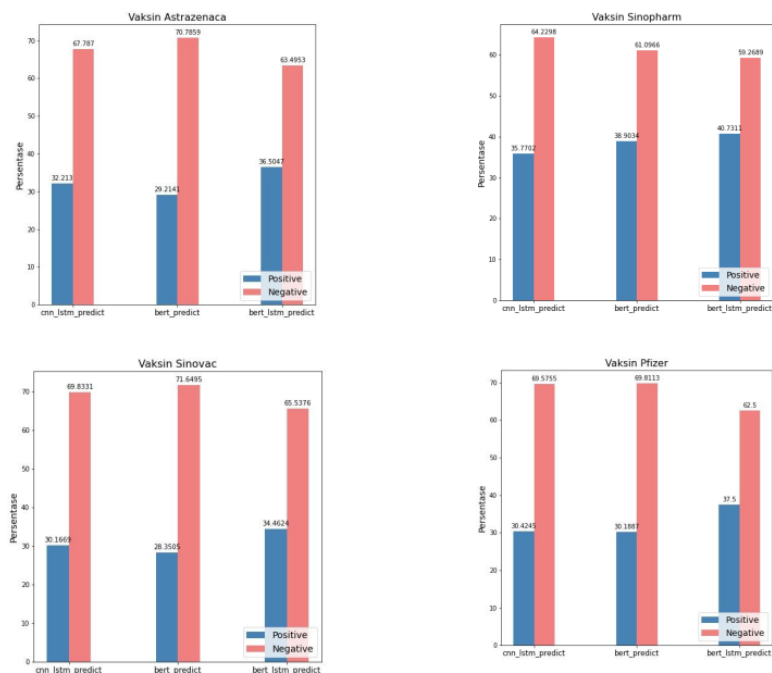
3.3. Opini Masyarakat Indonesia Terkait Jenis Vaksin COVID-19 di Indonesia

Saat ini terdapat beberapa jenis vaksin yang telah digunakan oleh fasilitas kesehatan di Indonesia seperti Sinovac, AstraZeneca, Sinopharm, Pfizer [16]. Pada umumnya ketika banyak pilihan masyarakat akan dibingungkan untuk memilih jenis vaksin yang akan digunakan. Hal ini menyebabkan masyarakat mengeluarkan opininya terkait keberagaman jenis vaksin COVID-19. Opini yang digunakan untuk dilakukan analisis sentimen di crawling pada aplikasi twitter dengan menggunakan keyword: "vaksin", "vaksinasi", "vaksinisasi", "covid", "corona", "covid-19", "covid19", "sinovac", "astrazeneca", "sinopharm", "pfizer" dengan contoh hasil sebagai berikut :

Tabel 3.3 Sampel Hasil Prediksi Model terhadap Jenis Vaksin COVID-19

Kalimat	Prediksi BERT	Prediksi CNN - LSTM	Prediksi BERT-LSTM
yg udah vaksin pake astrazeneca aman	1	1	1
adanya vaksin covid pfizer kita untuk syukur alhamdulillah	1	1	1
'tolong badan gue panas dingin pdhl baru vaksin sinovac az'	-1	-1	-1

Setelah dilakukan prediksi kategori, dapat dilakukan pengelompokan sesuai hasil kategorinya dan didapatkan persentase sebagai berikut :



Gambar 3.3 Persentase Sentimen terhadap Jenis Vaksin

Dari gambar diatas terlihat bahwa masyarakat cenderung memiliki sentimen negatif terhadap seluruh jenis vaksin. Hal tersebut, dapat terlihat dari persentase vaksin Sinovac yang dihasilkan oleh model CNN-LSTM 70% negatif dan 30% positif, sedangkan model BERT 72% negatif dan 29% positif, dan BERT-LSTM 26% negatif dan 74% positif dari total 4074 dataset. Vaksin Pfizer dengan model CNN-LSTM 70% negatif dan 30% positif, model BERT 70% negatif dan 30% positif, dan BERT-LSTM 62,5% negatif dan 37,5% positif dari total 1272 dataset. Vaksin AstraZeneca dengan model CNN-LSTM 68% negatif dan 32% positif, model BERT 71% negatif dan 29% positif, dan BERT-LSTM 63% negatif dan 37% positif dari total 1934 dataset. Vaksin Sinopharm dengan model CNN-LSTM 64% negatif dan 36% positif, model BERT 61% negatif dan 39% positif, dan BERT-LSTM 59% negatif dan 41% positif dari total 383 dataset. Hal ini menunjukkan bahwa model CNN-LSTM, BERT, dan BERT-LSTM dengan menggunakan parameter pada tabel 2.1 menunjukkan hasil yang serupa kasus ini.

4. KESIMPULAN

Sejak 2019 hingga saat ini pandemi virus COVID-19 belum kunjung berakhir. Bahkan untuk saat ini virus COVID-19 sudah bermutasi hingga varian Delta, yang dimana Delta menyebabkan lonjakan angka positif dan kematian yang tinggi di seratus negara lebih yang ada di dunia termasuk Indonesia. Dalam penekanan angka positif dan kematian di Indonesia sudah menerapkan beberapa upaya seperti pembatasan sosial, edukasi kebersihan, vaksinasi dan masih banyak lainnya. Namun untuk saat ini vaksin merupakan salah satu jalan keluar yang terbaik. Opini publik terhadap vaksin COVID-19 merupakan hal yang penting, dimana hal ini dapat menjadi gambaran untuk memahami kondisi masyarakat Indonesia terkait vaksinasi COVID-19. Opini masyarakat biasanya diungkapkan melalui media sosial Twitter yang memiliki sekitar 19.5 juta pengguna dan menjadi salah satu aplikasi paling sering digunakan oleh masyarakat Indonesia, sehingga posting pada Twitter dapat jadi sebagai bahan data untuk melakukan analisis sentimen terkait opini publik terhadap vaksinasi COVID-19.

Berdasarkan hasil analisis data postingan pada Twitter dari 19 Januari 2021 hingga 19 Juli 2021 masyarakat cenderung memiliki sentimen negatif. Dibuktikan pada hasil analisis opini publik terkait keefektifan vaksin pada setiap model dengan angka persentase sentimen negatif: BERT 70% , CNN-LSTM 58%, BERT-LSTM 67% dari total 8411 dataset. Pada opini publik terkait vaksin berbayar menghasilkan persentase sentimen negatif: BERT 77% , CNN-LSTM 65%, BERT-LSTM 74% dari total 5177 dataset, sedangkan vaksin gratis menghasilkan persentase sentimen negatif : BERT 61% , CNN-LSTM 7%, BERT-LSTM 58% dari total 5483 dataset. Terkait opini publik pada jenis vaksin di Indonesia menghasilkan persentase sentimen negatif untuk jenis vaksin Pfizer: BERT 70% , CNN-LSTM 70%, BERT-LSTM 62.5% dari total 1272 dataset, Sinovac memiliki nilai persentase negatif: BERT 72% , CNN-LSTM 70%, BERT-LSTM 66% dari total 4074 dataset, Sinopharm memiliki nilai persentase sentimen negatif: BERT 64% , CNN-LSTM 61%, BERT-LSTM 59% dari total 383 dataset, dan AstraZeneca memiliki persentase sentimen negatif: BERT 71% , CNN-LSTM 68%, BERT-LSTM 63% dari total 1934 dataset. Jika dilihat dari hasil evaluasi model, BERT memiliki kemampuan yang lebih baik jika dibandingkan dengan model CNN-LSTM, karena model BERT memiliki *accuracy, loss, precision, recall, F1* yang lebih baik.

Dari hasil analisis sentimen terhadap opini publik terkait vaksinasi COVID-19 di Indonesia yang lebih cenderung negatif perlu upaya dari pemerintah, masyarakat dan pribadi masing - masing untuk membantu membantu dalam edukasi vaksin COVID-19. Karena untuk saat ini vaksinasi merupakan salah satu jalan keluar terbaik untuk menciptakan kekebalan tubuh pada setiap masyarakat dengan harapan dapat membawa Indonesia menjadi lebih maju dengan terbebas dari pandemi COVID-19.

REFERENSI

- [1] WHO 2021, WHO Director-General's opening remarks at the media briefing on COVID-19 – 12 July 2021, Available at: <https://www.who.int/director-general/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19-12-july-2021>.
- [2] Y. Liu, A. A. Gayle, A. Wilder-Smith, and J. Rocklöv 2020, "The reproductive number of COVID-19 is higher compared to SARS coronavirus," J. Travel Med., vol. 27, no. 2, pp. 1–4, Mar. 2020, doi: 10.1093/jtm/taaa021.
- [3] Valerisha, A 2020, 'Pandemi Global COVID-19 dan Problematika Negara-Bangsa: Transparansi Data Sebagai Vaksin Socio-digital?', Jurnal Ilmiah Hubungan Internasional, Edisi Khusus.
- [4] Perbawa, I 2021, "KEBIJAKAN PEMERINTAH INDONESIA DALAM MENANGGULANGI COVID-19 BERDASARKAN INSTRUMEN HUKUM INTERNASIONAL" ,Jurnal Ilmu Sosial dan Humaniora, Vol. 10, No. 1, April 2021, 199-200.
- [5] T. T. Mengistie and D. Kumar, "Deep Learning Based Sentiment Analysis On COVID-19 Public Reviews," 2021 International Conference on Artificial Intelligence in Information and Communication (ICAIIIC), 2021, pp. 444-449, doi: 10.1109/ICAIIIC51459.2021.9415191.
- [6] R. Kaur and S. Ranjan, "Sentiment analysis of 21 dayscovid-19 indian lockdown tweets," International Journal of Advance Research in Science and Engineering, vol. 9, no. 7, pp. 37-44, 2020.
- [7] Z. Gao, A. Feng, X. Song, and X. Wu, "Target-dependent sentiment classification with BERT," IEEE Access, vol. 7, pp. 154290–154299, 2019.
- [8] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language Understanding," in Proc. NAACL-HLT, 2019, pp. 4171–4186.
- [9] D. I. Komputer, F. Matematika, D. A. N. Ilmu, and P. Alam, "Prediksi temporal untuk kemunculan titik panas di kabupaten rokan hilir riau menggunakan long short term memory rnn hafshah luthfiah," 2018.
- [10] L. Wiranda dan M. Sadikin 2019, "PENERAPAN LONG SHORT TERM MEMORY PADA DATA TIME SERIES UNTUK MEMPREDIKSI PENJUALAN PRODUK PT. METISKA FARMA" , Jurnal Nasional Pendidikan Teknik Informatika, Volume 8, Nomor 3, Desember 2019.
- [11] Fajri Koto, Afshin, Rahimi, Jey Han Lau, and Timothy Baldwin. IndoLEM and IndoBERT: A Benchmark Dataset and Pre-trained Language Model for Indonesian NLP. In Proceedings of the 28th COLING, December 2020.
- [12] S. Kannan and V. Gurusamy, "Preprocessing techniques for text mining,"International Journal of Computer Science & Communication Networks, vol. 5, no. 1, pp. 7-16, 2014.
- [13] A. A. Maarif, "Penerapan Algoritma TF-IDF untuk Pencarian Karya Ilmiah", Udinus Repo, 2015.
- [14] R. R. A. Siregar, F. A. Sinaga, R. Arianto, 'APLIKASI PENENTUAN DOSEN PENGUJI SKRIPSI MENGGUNAKAN METODE TF-IDF DAN VECTOR SPACE MODEL', Journal of Computer Science and Information Systems, volume 1, no 2, hal 171-186, Oktober 2017.
- [15] Kemdikbud.go.id. Hasil Pencarian - KBBI Daring. [online] Available at: <https://kbbi.kemdikbud.go.id/entri/keefektifan>, 2016.
- [16] Wadrianto, G.K., Kenali, 6 Jenis Vaksin Covid-19 yang Dipakai di Indonesia Halaman all - Kompas.com. KOMPAS.com. Available at: <https://lifestyle.kompas.com/read/2021/07/15/155004420/kenali-6-jenis-vaksin-covid-19-yang-dipakai-di-indonesia?page=all>, 2021.
- [17] B. Wilie, K. Vincentio, G. I. Winata, S. Cahyawijaya, X. Li, Z. Y. L. ,S. Soleman, R. Mahendra, P. Fung, S. Bahar and A. Purwarianti, 'IndoNLU: Benchmark and Resources for Evaluating Indonesian Natural Language Understanding', Proceedings of the 1st Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 10th International Joint Conference on Natural Language Processing, 2020.
- [18] Purwarianti, A. and I. A. P. A. Crisdayanti, 'Improving Bi-LSTM Performance for Indonesian Sentiment Analysis Using Paragraph Vector', Proceedings of the 2019 International Conference of Advanced Informatics: Concepts, Theory and Applications (ICAICTA), pages 1-5, IEEE 2019.
- [19] A. Yenter and A. Verma, "Deep CNN-LSTM with combined kernels from multiple branches for IMDb review sentiment analysis," 2017 IEEE 8th Annual Ubiquitous Computing, Electronics and Mobile Communication Conference (UEMCON), 2017, pp. 540-546, doi: 10.1109/UEMCON.2017.8249013.
- [20] C. Arumsari, E. Yulianto, and E. Nur' Afifah, "SOSIALISASI DALAM RANGKA MEMELIHARA KESADARAN WARGA PADA KESEHATAN DI MASA PANDEMI COVID-19", .bernas jurnal .pengabdian .masyarakat, vol. 2, no. 1, pp. 272-276, Jan. 2021.
- [21] A. Ristyawati, "Efektivitas Kebijakan Pembatasan Sosial Berskala Besar Dalam Masa Pandemi Corona Virus 2019 oleh Pemerintah Sesuai Amanat UUD NRI Tahun 1945," Administrative Law and Governance Journal, vol. 3, no. 2, pp. 240-249, Jun. 2020. <https://doi.org/10.14710/alj.v3i2.240-249>

Gemastik XIV: Yaudahlah

ORIGINALITY REPORT

19%

SIMILARITY INDEX

18%

INTERNET SOURCES

13%

PUBLICATIONS

10%

STUDENT PAPERS

PRIMARY SOURCES

1

arxiv.org

Internet Source

2%

2

Tajebe Tsega Mengistie, Deepak Kumar.
"Deep Learning Based Sentiment Analysis On
COVID-19 Public Reviews", 2021 International
Conference on Artificial Intelligence in
Information and Communication (ICAIIIC),
2021

Publication

1%

3

jurnal.mdp.ac.id

Internet Source

1%

4

Lorenzo Pellis, Francesca Scarabel, Helena B
Stage, Christopher E Overton et al.
"Challenges in control of Covid-19: short
doubling time and long delay to effect of
interventions", Cold Spring Harbor
Laboratory, 2020

Publication

1%

5

ejournal2.undip.ac.id

Internet Source

1%

6	arrow.tudublin.ie Internet Source	1 %
7	github.com Internet Source	1 %
8	Submitted to Sekolah Ciiputra High School Student Paper	1 %
9	publikasiilmiah.unwahas.ac.id Internet Source	1 %
10	www.hindawi.com Internet Source	1 %
11	ejournal.ihtdn.ac.id Internet Source	1 %
12	Submitted to University of Melbourne Student Paper	<1 %
13	journal.unpar.ac.id Internet Source	<1 %
14	Submitted to Universitas Raharja Student Paper	<1 %
15	repository.ipb.ac.id:8080 Internet Source	<1 %
16	Submitted to Binus University International Student Paper	<1 %
17	Warih Maharani. "Sentiment Analysis during Jakarta Flood for Emergency Responses and	<1 %

Situational Awareness in Disaster Management using BERT", 2020 8th International Conference on Information and Communication Technology (ICoICT), 2020

Publication

18

ejournal.undiksha.ac.id

Internet Source

<1 %

19

hal.inria.fr

Internet Source

<1 %

20

Inswasti Cahyani, Eko W Putro, Asep M Ridwanuloh, Satrio Wibowo et al. "Genome Profiling of SARS-CoV-2 in Indonesia, ASEAN and the Neighbouring East Asian Countries: Features, Challenges and Achievements", Cold Spring Harbor Laboratory, 2021

Publication

<1 %

21

media.neliti.com

Internet Source

<1 %

22

Tianyi Wang, Ke Lu, Kam Pui Chow, Qing Zhu. "COVID-19 Sensing: Negative Sentiment Analysis on Social Media in China via BERT Model", IEEE Access, 2020

Publication

<1 %

23

academicimmftui.blogspot.com

Internet Source

<1 %

24

es.scribd.com

Internet Source

<1 %

25	www.medcom.id Internet Source	<1 %
26	www.researchgate.net Internet Source	<1 %
27	blog.heylaw.id Internet Source	<1 %
28	repository.dl.itc.u-tokyo.ac.jp Internet Source	<1 %
29	text-id.123dok.com Internet Source	<1 %
30	www.kompasiana.com Internet Source	<1 %
31	www.sv-jme.eu Internet Source	<1 %
32	gosocial.co.id Internet Source	<1 %
33	positori.usu.ac.id Internet Source	<1 %
34	www.slideshare.net Internet Source	<1 %
35	Muhamad Hardi, Muhammad Nur Firdaus, Bayu Putra Pamungkas, Usman Sudibyo et al. "Leaf morphological feature extraction based on K-Nearest Neighbor", 2018 International	<1 %

Conference on Information and Communications Technology (ICOIACT), 2018

Publication

36

impormasi.wordpress.com

Internet Source

<1 %

37

library.binus.ac.id

Internet Source

<1 %

38

repository.its.ac.id

Internet Source

<1 %

39

www.idntimes.com

Internet Source

<1 %

40

Evan Kusuma Susanto, Yosi Kristian.
"Pemanfaatan Asynchronous Advantage
Actor-Critic Dalam Pembuatan AI Game Bot
Pada Game Arcade", Journal of Intelligent
System and Computation, 2019

Publication

<1 %

Exclude quotes

Off

Exclude matches

Off

Exclude bibliography

On

Gemastik XIV: Yaudahlah

PAGE 1

PAGE 2

PAGE 3

PAGE 4

PAGE 5

PAGE 6

PAGE 7

PAGE 8

PAGE 9

PAGE 10