

Deteksi Berita *Online Hoax* Covid-19 Di Indonesia Menggunakan Metode *Hybrid Long Short-Term Memory* dan *Support Vector Machine*

Dwi Fitriani Nur Anisa, Mohammad Iqbal, dan Imam Mukhlash
Departemen Matematika, Institut Teknologi Sepuluh Nopember (ITS)
e-mail: iqbal@matematika.its.ac.id; imamm@matematika.its.ac.id.

Abstrak—Fokus masyarakat Indonesia tidak lepas dari kasus pandemi *Coronavirus Disease 2019 (COVID-19)* dengan mengikuti setiap informasi terkait perkembangannya setiap hari. Hal ini yang mendorong banyak pihak terlebih pemerintah untuk menyediakan layanan informasi terkini terkait COVID-19. Namun, banyak berita *online* menyajikan informasi palsu yang dikenal dengan berita *hoax* tentang COVID-19 yang dapat menyebabkan keresahan masyarakat. Pada Tugas Akhir ini, dilakukan deteksi terhadap berita-berita *online* seputar informasi COVID-19 di Indonesia yang dibagi menjadi dua kategori, yaitu berita *hoax* dan berita fakta. Proses deteksi berita *online* dilakukan dengan metode penggabungan *Long-Short Term Memory* dan *Support Vector Machine (hybrid LSTM-SVM)*. LSTM menghasilkan fitur teks representatif yang selanjutnya digunakan untuk proses klasifikasi berita oleh SVM yang menghasilkan persentase nilai akurasi mencapai 94%. Nilai tersebut lebih tinggi dibandingkan dengan hanya mengimplementasikan Metode LSTM atau Metode SVM saja.

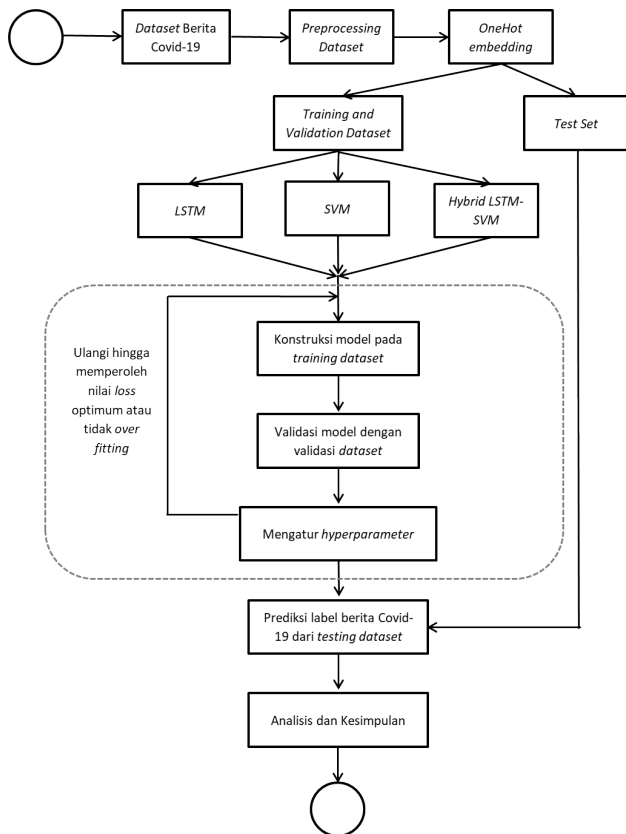
Kata Kunci—COVID-19, Deteksi Berita *Hoax*, LSTM, SVM.

I. PENDAHULUAN

PADA penghujung akhir tahun 2019, hampir seluruh negara di dunia digemparkan dengan penyebaran virus influenza baru yang disebut dengan *Coronavirus Disease 2019*, kemudian dikenal dengan COVID-19. Awal mula penemuan penyakit COVID-19 yaitu di Wuhan, Negara Cina dengan kasus 5 orang terjangkit pada sekitar 18 – 29 Desember 2019 [1]. Dengan tingkat penyebaran tinggi, kasus terinfeksi COVID-19 telah terjadi hampir di seluruh negara di awal tahun 2020, termasuk Indonesia. Sejak 2 Maret 2020, tercatat telah ditemukan kasus terinfeksi COVID-19 di Indonesia, lebih tepatnya di kota Jakarta, untuk pertama kalinya dengan jumlah penderita sebanyak 2 orang[2]. Sebagai catatan pengingat bahwa kasus terinfeksi COVID-19 di Indonesia sampai 31 Mei 2021 menembus angka kumulatif 1.816.041. Setelah kurang lebih dua tahun lamanya, kasus terinfeksi COVID-19 masih ditemukan yang ditambah dengan penemuan varian baru dari COVID-19 seperti varian delta, varian omicron, dan sebagainya[3]. Dengan mempertimbangkan jejak rekam kasus penyebaran COVID-19 di Indonesia yang semakin meningkat semenjak bulan Maret 2020, pemerintah Indonesia segera membentuk suatu tim yang dikenal dengan Satuan Tugas (Satgas) Penanganan COVID-19 dan dibagi ke dalam dua kelompok, yaitu Satgas COVID-19 dan Satgas Pemulihan Ekonomi Nasional di bawah pengawasan Komite Pemulihan Nasional. Salah satu

tugas dari Satgas COVID-19, yaitu untuk melaporkan data terkait COVID-19 di Indonesia kepada seluruh masyarakat Indonesia agar mendapatkan perkembangan informasi terkini terkait penyebaran COVID-19 di Indonesia serta langkah pencegahan atau penanggulangannya. Pemberian informasi tersebut disebarluaskan melalui beberapa laman resmi pemerintahan, seperti <https://covid19.go.id/>, <https://kawal covid19.id/> dan lain sebagainya. Akan tetapi, berita-berita yang memberikan informasi palsu tentang COVID-19 banyak bermunculan sehingga mampu meresahkan dan mengakibatkan salah interpretasi dalam melawan pandemi COVID-19. Berita tersebut yang selanjutnya lebih dikenal dengan berita *hoax*.

Pada penelitian ini, dilakukan deteksi terhadap berita *hoax* tentang COVID-19 di Indonesia sehingga dapat meminimalisir atau menghilangkan berita-berita *hoax* COVID-19 di Indonesia. Terdapat beberapa penelitian dalam mendeteksi berita *hoax* menggunakan berbagai metode, antara lain: Prannay S. Reddy dkk (2019) melakukan penelitian untuk mengidentifikasi berita palsu secara akurat dengan menerapkan teknik *natural language processing* dan *machine learning* serta menyelidiki dan membandingkan akurasi teknik Naïve Bayes, SVM, jaringan syaraf dan LSTM untuk menemukan metode yang paling cocok. Data yang digunakan pada penelitian ini berupa berita tentang politik Amerika Serikat yang diperoleh dari Kaggle sebanyak kurang lebih 33.000 artikel data. Langkah pertama yang dilakukan adalah *pre-processing* data. Kemudian, dilanjutkan dengan memprediksi *class* artikel berita yang diberikan menjadi nyata atau palsu. Data dipartisi menjadi train yang memiliki 26.000 korpus artikel berita dan test yang memiliki 7.000 korpus artikel berita. Selanjutnya, akurasi keempat metode tersebut dibandingkan dan diperoleh persentase 94% untuk LSTM sebagai skor akurasi tertinggi[4]. Aditya Gusti Tamam dkk (2018) melakukan penelitian pendeteksian berita *hoax* pada media sosial dengan *text mining* berbasis sistem klarifikasi. Sistem pada penelitian ini dibangun menggunakan Algoritma Peringkat Teks untuk ekstraksi kata kunci dan Algoritma *Cosine Similarity* untuk menghitung tingkat kemiripan dokumen. Hasil ekstraksi kata kunci digunakan untuk mencari konten terkait menggunakan masukan dari pengguna melalui mesin pencari, kemudian menghitung nilai kemiripan. Jika konten yang terkait cenderung berasal dari media terpercaya, maka konten tersebut berpotensi faktual. Jika konten terkait cenderung



Gambar 1. Alur metode deteksi berita *online hoax* COVID-19 di Indonesia.

dipublikasikan oleh media yang tidak dapat diandalkan, maka ada potensi sebuah tipuan[5].

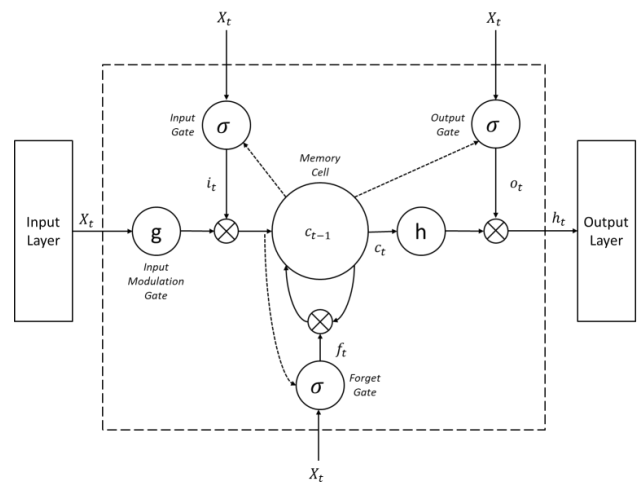
Penelitian sebelumnya memerlukan tahapan ekstraksi fitur yang membutuhkan waktu dan pengetahuan dari seorang ahli. Untuk menghindari hal tersebut, penelitian ini mengimplementasikan metode penggabungan antara *Long Short Term Memory* (LSTM) dan *Support Vector Machines* (SVM) yang disebut dengan *hybrid LSTM-SVM*. Penggabungan kedua metode tersebut dipilih karena memiliki tingkat akurasi yang tinggi dan dapat meningkatkan kinerja model LSTM dan performa model SVM[6]. Secara umum, LSTM digunakan untuk membangkitkan fitur representatif dan dilanjutkan dengan metode SVM untuk mendeteksi berita *hoax* dari hasil representasi fitur yang telah dibangkitkan. Alur metode deteksi berita *online hoax* COVID-19 di Indonesia dapat dilihat pada Gambar 1.

II. METODE PENELITIAN

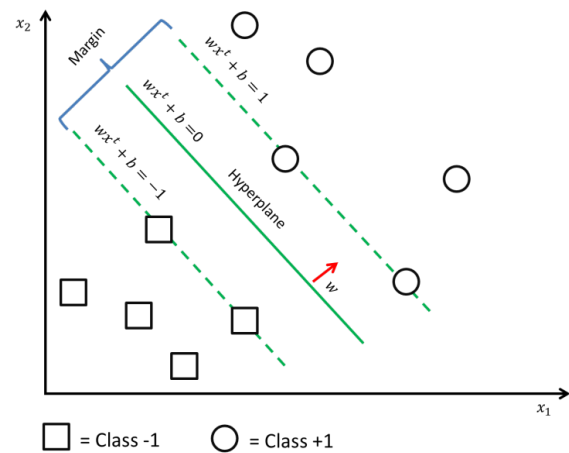
Penelitian ini mengimplementasikan metode gabungan antara *Long-Short Term Memory* (LSTM) dan *Support Vector Machines* (SVM) untuk memperoleh hasil analisis dari klasifikasi berita mengenai COVID-19 di Indonesia. Metode yang digunakan memiliki beberapa tahap, yaitu pengumpulan berita, *preprocessing* data yang mencakup pembersihan data (*cleansing data*) dan mengubah data ke bentuk frekuensi, membagi data menjadi data *train* dan data *testing*, pembentukan metode, pengujian metode serta analisis akurasi dan berita *online hoax* COVID-19 di Indonesia.

A. Long-Short Term Memory (LSTM)

Jaringan *Long-Short Term Memory* (LSTM) merupakan salah satu tipe arsitektur jaringan dari *Recursive Neural*



Gambar 2. Desain unit memori LSTM.



Gambar 3. Model SVM linier.

Network (RNN) yang mampu mempelajari relasi jangka panjang antar data. LSTM juga dapat memecahkan masalah gradien yang menghilang, yaitu saturasi neuron lapisan sebelumnya saat informasi masuk lebih dalam ke jaringan. Artinya, neuron yang menyimpan informasi berharga tetapi terletak pada titik waktu yang jauh, berhenti berkontribusi ke jaringan yang mengakibatkan training yang tidak efisien[7].

Pada LSTM, terdapat kekhasan *cell* yang dikonfigurasi oleh empat buah gerbang utama: *input gate*, *input modulation gate*, *forget gate*, dan *output gate*. *Input gate* mengambil poin *input* baru dari luar dan proses data yang baru. *Memory cell input gate* mengambil *input* dari *output cell* LSTM pada iterasi terakhir. *Forget gate* memutuskan kapan hasil *output* dilupakan dan dengan demikian memilih jeda waktu optimal untuk urutan *input*. *Output gate* mengambil seluruh hasil perhitungan dan menghasilkan *output* dari *cell* LSTM[8]. Proses tersebut diilustrasikan pada Gambar 2.

Keadaan tersembunyi dari *memory cells* ditunjukkan oleh perhitungan formula berikut[8]:

- fungsi *input gate*

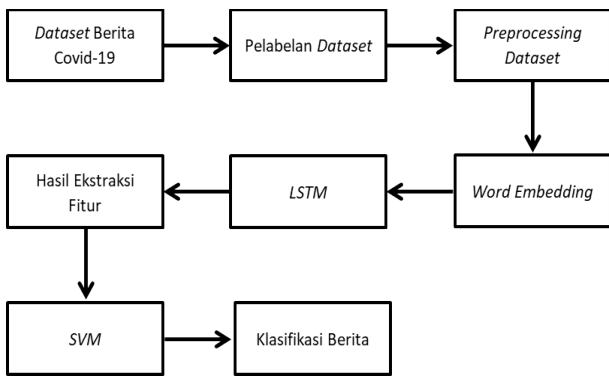
$$i_t = \sigma(W_{ix}x_t + W_{ih}h_{t-1} + W_{ic}c_{t-1} + b_i) \quad (1)$$

- fungsi *forget gate*

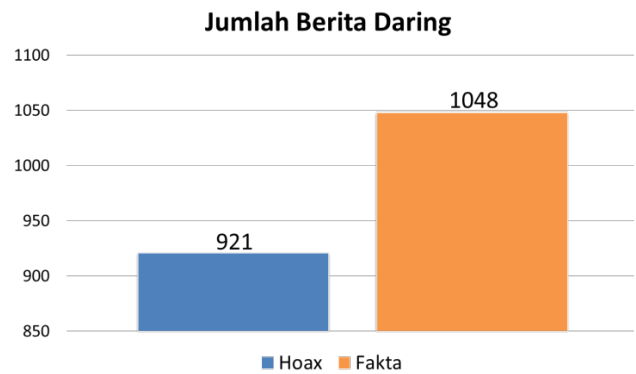
$$f_t = \sigma(W_{fx}x_t + W_{fh}h_{t-1} + W_{fc}c_{t-1} + b_f) \quad (2)$$

- fungsi *memory cell*

$$c_t = f_t * c_{t-1} + i_t * g(W_{cx}x_t + W_{ch}h_{t-1} + W_{cc}c_{t-1} + b_c) \quad (3)$$



Gambar 4. Arsitektur hybrid LSTM-SVM.



Gambar 5. Banyaknya data berita online tentang COVID-19 di Indonesia dengan label hoax dan fakta

Tabel 1.
Data berita online tentang COVID-19 di Indonesia

No	Date	Headline	Label
1	2020-04-22	Dari Ratusan Ribu Tahanan Tidak Ada Satupun Aktivis Islam yang Dibebaskan di Tengah Wabah COVID-19	Hoax
2	2020-09-07	Ganjil Genap Di Jakarta Ditiadakan Mulai Senin, 7 September 2020	Hoax
3	2021-05-12	Analisis Data Covid-19 Indonesia (Update Per 09 Mei 2021)	Fakta
4	2021-01-01	Keputusan Menteri Kesehatan Republik Indonesia Nomor Hk.01.07/Menkes/12757/2020	Fakta
5	2020-11-06	Pasien Sembuh Kumulatif Mencapai 360.705 Orang	Fakta
...
1969	2021-03-12	Pasien Sembuh Meningkatkan Lagi Mencapai 1.231.454 Orang	Fakta

Tabel 2.
Case folding

Sebelum	Sesudah
Imigran China Sengaja Membawa Virus, Tujuannya Membasmi Rakyat +62	imigran china sengaja membawa virus, tujuannya membasmi rakyat +62
Kesembuhan COVID-19 Terus Meningkatkan Mencapai 1.426.145 Orang	kesembuhan covid-19 terus meningkat mencapai 1.426.145 orang

Tabel 3.
Tokenizing

Sebelum	Sesudah
imigran china sengaja membawa virus, tujuannya membasmi rakyat +62	imigran china sengaja membawa virus tujuannya membasmi rakyat 62
kesembuhan covid-19 terus meningkat mencapai 1.426.145 orang	kesembuhan covid19 terus meningkat mencapai 1426145 orang

- fungsi *output gate*

$$o_t = \sigma(W_{ox}x_t + W_{oh}h_{t-1} + W_{oc}c_{t-1} + b_o) \tag{4}$$

dimana, W melambangkan matriks bobot, b melambangkan vektor bias, g dan h melambangkan perpanjangan dari fungsi sigmoid dengan rentang perubahan ke $[-2, 2]$ dan $[-1, 1]$, $*$ melambangkan produk skalar dua vektor atau matriks dan σ merupakan singkatan fungsi sigmoid yang ditetapkan dalam Pers. 5[8].

$$\sigma(x) = \frac{1}{1+e^x} \tag{5}$$

Untuk fungsi objektif digunakan fungsi *loss* kuadrat yang ditunjukkan oleh formula berikut[8]:

$$\sigma(x) = \sum_{t=1}^n (y_t - p_t)^2 \tag{6}$$

dimana, y melambangkan *output* nyata dan p melambangkan prediksi berita.

B. Support Vector Machine (SVM)

Pada awalnya, algoritma *Support Vector Machines* (SVM) diusulkan untuk membangun pengklasifikasian linier pada tahun 1963 oleh Vapnik. Hal ini bertujuan untuk membentuk batas keputusan antara dua *class* yang memungkinkan prediksi label dari satu atau lebih vektor fitur[9]. SVM merupakan metode *learning machines* yang bekerja atas prinsip *Structural Risk Minimization* (SRM) dengan tujuan

untuk menemukan *hyperplane* terbaik yang memisahkan dua buah *class* pada *input space*[10].

Gambar 3 memperlihatkan beberapa *pattern* yang merupakan anggota dari dua buah *class*, yaitu *pattern* pada *class* -1 disimbolkan dengan bentuk persegi sedangkan *pattern* Gambar 4. Arsitektur hybrid LSTM-SVM pada *class* +1 disimbolkan dengan bentuk lingkaran. Proses pembelajaran dalam masalah klasifikasi yang dapat diterjemahkan sebagai upaya untuk menemukan garis (*hyperplane*) optimal yang memisahkan antara kedua *class* tersebut. Batas keputusan yang dikenal sebagai *hyperplane* diorientasikan sedemikian rupa sehingga sejauh mungkin dari titik data terdekat dari masing-masing *class*. Titik terdekat ini disebut *support vectors*[10].

Diberikan set data pelatihan dengan label [10]:

$$(x_1, y_1), \dots, (x_n, y_n), x_i \in R^d \text{ dan } y_i \in (-1, +1) \tag{7}$$

dimana x_i merupakan representasi vektor fitur dan y_i merupakan label *class* (negatif atau positif) dari kompleks *training* i .

Hyperplane optimal dapat didefinisikan sebagai berikut[9]:

$$wx^t + b = 0 \tag{8}$$

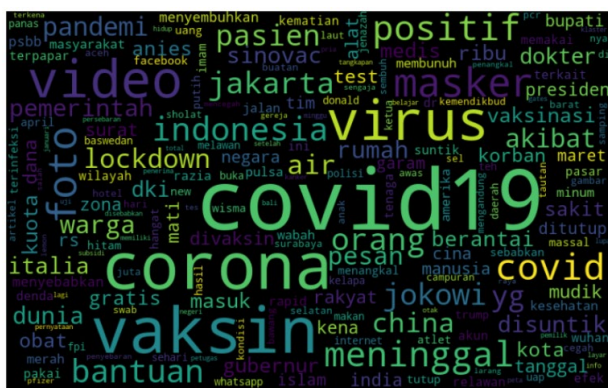
dimana w merupakan bobot vektor, x merupakan vektor fitur masukan dan b merupakan biasnya. w dan b akan memenuhi ketidaksamaan berikut untuk semua elemen set pelatihan[9]:

Tabel 4.
Filtering

Sebelum	Sesudah
imigran china sengaja membawa virus tujuannya membasmi rakyat 62	imigran china sengaja membawa virus tujuannya membasmi rakyat 62
kesembuhan covid19 terus meningkat mencapai 1426145 orang	kesembuhan covid19 meningkat mencapai 1426145 orang

Tabel 5.
Stemming

Sebelum	Sesudah
imigran china sengaja membawa virus tujuannya membasmi rakyat 62	"imigran", "china", "sengaja", "bawa", "virus", "tuju", "basmi", "rakyat"
kesembuhan covid19 meningkat mencapai 1426145 orang	"sembuh", "tingkat", "capai", "orang"



Gambar 6. WordCloud berita online tentang COVID-19 di Indonesia dengan label hoax.

$$wx_i^t + b \geq +1, \text{ jika } y_i = +1 \tag{9}$$

$$wx_i^t + b \leq -1, \text{ jika } y_i = -11 \tag{10}$$

Tujuan dari melatih model SVM adalah untuk menemukan w dan b sehingga *hyperplane* memisahkan data dan memaksimalkan margin $\frac{1}{\|w\|^2}$.

C. Hybrid Long Short Term Memory dan Support Vector Machine (Hybrid LSTM-SVM)

Support Vector Machine (SVM) dikenal sebagai algoritma pembelajaran yang sangat efisien, namun algoritma ini menangkap fitur dengan minim dan berlainan dalam mengklasifikasi dokumen. Hal ini menjadi penyebab sulitnya mendeteksi hubungan dalam kalimat[11].

Long Short Term Memory (LSTM) mampu menangkap ketergantungan jangka panjang dalam sebuah kalimat. Jenis jaringan saraf ini memiliki keunggulan dalam mengklasifikasi beberapa tugas analisis sentimen sehubungan dengan algoritma pembelajaran yang umum digunakan[11].

Berdasarkan penjabaran mengenai LSTM dan SVM di atas, dapat diterapkan penggabungan antara kedua metode tersebut untuk mengambil keuntungan dari dua strategi klasifikasi. Arsitektur hybrid LSTM-SVM ditunjukkan oleh Gambar 4.

Pada tahap *training*, LSTM dilatih dengan mempertimbangkan data *training* dan label yang sesuai. Setelah model statistik LSTM dihitung, *embedding* data



Gambar 7. WordCloud berita online tentang COVID-19 di Indonesia dengan label fakta.

Tabel 6.
Hasil TF-IDF

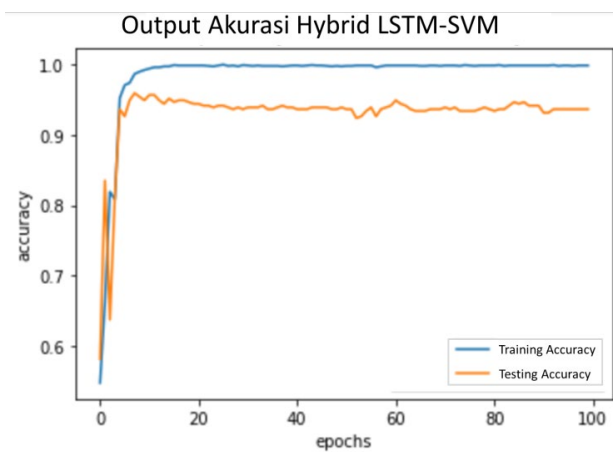
(Indeks Dokumen, Indeks Kata)	Nilai TF-IDF
(0, 4821)	0.3708399647590785
(0, 764)	0.3979937818159212
(0, 3434)	0.4195469789467403
(0, 3875)	0.4195469789467403
...	...
(2, 671)	0.29910835492758636
(2, 717)	0.32352910249292993
(2, 87)	0.32795931913186904
(3, 1913)	0.45441785384706573
(3, 957)	0.39116538894072317
...	...
(1964, 3687)	0.534397695734636
(1964, 4605)	0.3784934720448085
(1965, 965)	0.456776692280458
(1965, 3386)	0.456776692280458
...	...
(1968, 1821)	0.3779644730092272
(1968, 4803)	0.3779644730092272
(1968, 1966)	0.3779644730092272
(1968, 2001)	0.3779644730092272
(1968, 3849)	0.3779644730092272

dihitung dengan memanfaatkan bobot yang diperoleh dari *layer* sebelum *softmax classifier* (*layer* kedua dari belakang) dengan memberikan *input* dokumen yang dipertimbangkan pada LSTM untuk setiap data *training*. *Embedding* data digunakan sebagai fitur selama fase *training* pengklasifikasian SVM bersama dengan serangkaian fitur klasifikasi data yang banyak digunakan[11].

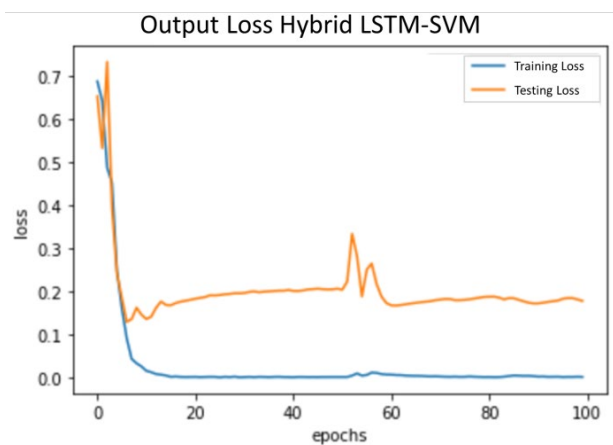
Pada tahap *testing*, untuk setiap data yang harus diklasifikasikan, vektor *embedding* diperoleh dengan memanfaatkan LSTM yang telah dilatih sebelumnya. *Embedding* digunakan bersama dengan fitur klasifikasi data lainnya oleh pengklasifikasian SVM yang menampilkan *class* yang diprediksi[11].

D. Pendeteksi Berita Online Hoax COVID-19

Hal pertama yang dilakukan dalam penelitian ini adalah pengumpulan data dari laman resmi pemerintahan, yaitu <https://covid19.go.id/> dan <https://kawalcovid19.id/>. Data yang diambil merupakan data berita tentang COVID-19 di Indonesia dari Maret 2020 hingga Mei 2021. Pada kedua



Gambar 8. Grafik akurasi *hybrid* LSTM-SVM.



Gambar 9. Grafik akurasi *loss* LSTM-SVM.

Tabel 7.
Tabel performansi metode

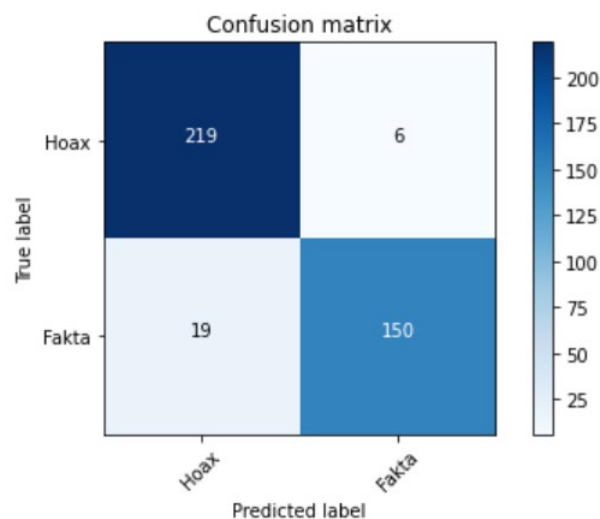
metode	accuracy	precision	recall	f1-score
SVM	76%	76%	75%	75%
LSTM	92%	92%	91%	92%
Hybrid LSTM-SVM	94%	94%	93%	93%

laman resmi pemerintahan tersebut, terdapat halaman yang menyajikan kumpulan berita-berita terkini tentang COVID-19, baik berita tentang COVID-19 dengan label *hoax* maupun berita tentang COVID-19 dengan label fakta.

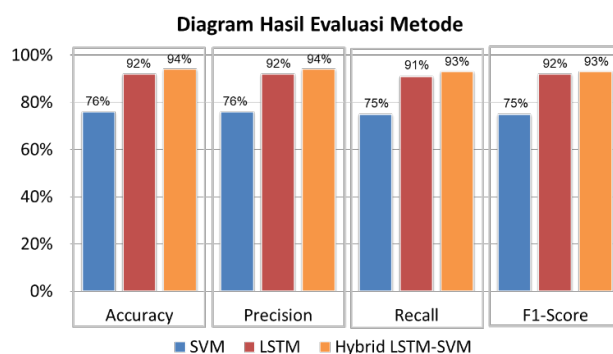
Kemudian dilanjutkan dengan seleksi pada setiap data untuk mendapatkan data yang diinginkan sebelum masuk ke tahap implementasi. *Cleansing* data dilakukan untuk membersihkan data berupa kalimat utuh agar dapat diproses ke tahap implementasi. Terdapat empat langkah dalam *cleansing* data, yaitu: *case folding*, *tokenizing*, *filtering*, dan *stemming*.

Selanjutnya, dilakukan proses konversi kata yang berupa karakter *alphanumeric* ke dalam bentuk vektor. Kemudian, data dibagi menjadi dua bagian, yaitu data *training* dan data *testing* dengan rasio yang telah ditentukan. Data *training* digunakan untuk melatih suatu algoritma dan data *testing* Gambar 1. Banyaknya data berita *online* tentang COVID-19 di Indonesia dengan label *hoax* dan fakta digunakan untuk menilai seberapa baik algoritma yang dilatih.

Selanjutnya, dilakukan prediksi label terhadap berita *online*, yaitu berita *online* tersebut tergolong ke dalam berita



Gambar 10. *Confusion matrix* *hybrid* LSTM-SVM.



Gambar 11. Diagram hasil evaluasi metode

hoax atau berita fakta dan juga untuk mengetahui seberapa akurat model yang digunakan dalam memprediksi berita *online* tersebut.

III. HASIL DAN PEMBAHASAN

Pada bagian ini dijelaskan tentang analisis hasil prediksi deteksi berita *hoax* COVID-19 di Indonesia dalam rentang waktu Maret 2020 hingga Mei 2021 dengan metode yang diusulkan yaitu, *hybrid* LSTM-SVM.

A. Pengumpulan Data Berita Online tentang COVID-19 di Indonesia

Pada tahap ini, data berita *online* tentang COVID-19 di Indonesia dikumpulkan dengan mengakses *website* resmi Satuan Tugas Penanganan COVID-19, yaitu <https://covid19.go.id/> dan <https://kawal covid19.id/>. Pengumpulan data berita *online* tentang COVID-19 di Indonesia dilakukan dengan bantuan *software* Microsoft Power BI Desktop. Dari kegiatan pengumpulan data berita *online* tentang COVID-19 di Indonesia tersebut, diperoleh 1969 data berita *online* tentang COVID-19 di Indonesia.

Selanjutnya, memberi label pada data berita *online* tentang COVID-19 di Indonesia. Pemberian label dilakukan dengan mengkategorikan data berita *online* tentang COVID-19 di Indonesia ke dalam dua kategori, yaitu berita *hoax* dan berita fakta. Hal ini seperti yang ditunjukkan pada Tabel 1.

Setelah dilakukan pelabelan pada data berita *online* tentang COVID-19 di Indonesia, dapat diketahui jumlah berita *online* tentang COVID-19 di Indonesia yang berlabelkan *hoax* dan

fakta, yaitu sebanyak 921 data berita *online* merupakan berita *hoax* dan sebanyak 1048 data berita *online* merupakan berita fakta seperti yang ditunjukkan pada Gambar 5.

B. Preprocessing Data Berita Online tentang COVID-19 di Indonesia

Pada tahap ini, *preprocessing* data dilakukan untuk memudahkan proses komputasi dalam mengenali data yang masih berupa kalimat utuh. Selanjutnya, dilakukan proses ekstraksi sentimen pada berita *online*. Tahap *preprocessing* data pada penelitian ini, yaitu:

- 1) *Case Folding*: untuk mengkonversi seluruh teks dalam suatu dokumen menjadi bentuk standar (diubah ke dalam bentuk huruf kecil). Hal ini dilakukan untuk menyamaratakan penggunaan huruf kapital karena data yang diperoleh tidak selalu konsisten dalam penggunaan huruf kapital. Perbedaan data berita *online* tentang COVID-19 di Indonesia sebelum dan sesudah melalui proses *case folding* ditunjukkan oleh Tabel 2.
- 2) *Tokenizing*: untuk memotong sekumpulan karakter dalam suatu teks ke dalam satuan kata dengan membedakan karakter tertentu yang dapat digunakan sebagai pemisah kata atau bukan. Perbedaan data berita *online* tentang COVID-19 di Indonesia sebelum dan sesudah melalui proses *tokenizing* ditunjukkan oleh .
- 3) *Filtering*: untuk mengambil kata penting dari hasil *tokenizing* dengan membuang kata yang kurang penting ataupun menyimpan kata yang penting. Perbedaan data berita *online* tentang COVID-19 di Indonesia sebelum dan sesudah melalui proses *filtering* ditunjukkan oleh Tabel 4.
- 4) *Stemming*: digunakan untuk memperkecil jumlah indeks yang berbeda dari sebuah dokumen dan untuk mengelompokkan kata yang memiliki kata dasar dan arti serupa namun memiliki bentuk berbeda karena mendapat imbuhan berbeda. Perbedaan data berita *online* tentang COVID-19 di Indonesia sebelum dan sesudah melalui proses *stemming* ditunjukkan oleh Tabel 5.

Setelah melalui empat tahap di atas, dapat diketahui kata apa saja yang paling sering muncul dengan mengimplementasikan WordCloud. Gambar 6 merupakan WordCloud untuk berita *online* tentang COVID-19 di Indonesia yang berlabel *hoax* dan Gambar 7 merupakan WordCloud untuk berita *online* tentang COVID-19 di Indonesia yang berlabel fakta. Dapat dilihat pada kedua WordCloud tersebut bahwa kata 'covid19' dan 'vaksin' memiliki ukuran yang paling besar. Hal ini menunjukkan bahwa kata 'covid19' dan 'vaksin' merupakan kata yang sering muncul dan merupakan topik yang banyak dibicarakan, baik pada kelompok berita *hoax* maupun berita fakta.

Pada kedua WordCloud tersebut juga dapat dilihat perbedaan kata yang muncul pada berita *hoax* dan berita fakta sebagai topik yang hangat dibicarakan pada kedua kelompok berita tersebut. Berdasarkan Gambar 6, kata-kata yang cenderung muncul pada berita *online* tentang COVID-19 di Indonesia yang berlabel *hoax* yaitu, video, foto, masker, positif, meninggal dan lain-lain sedangkan berdasarkan Gambar 7, kata-kata yang cenderung muncul pada berita *online* tentang COVID-19 di Indonesia yang berlabel fakta yaitu, pasien, data, sembuh dan lain-lain.

C. Mengubah Data Berita Online tentang COVID-19 di Indonesia ke Bentuk Frekuensi

Data berita *online* tentang COVID-19 di Indonesia diubah ke dalam bentuk frekuensi dengan menggunakan metode TF-IDF (*Term Frequency-Inverse Document Frequency*) ditunjukkan oleh Tabel 6. Hal ini bertujuan untuk memberi bobot terhadap kemunculan kata dalam suatu dokumen. Kata yang sering muncul di satu dokumen dan jarang muncul di dokumen lain akan memiliki bobot yang tinggi sedangkan kata yang sering muncul di banyak dokumen akan memiliki bobot yang lebih rendah. Semakin tinggi bobot kata tersebut, maka semakin penting pula kata tersebut dalam kumpulan dokumen.

D. Train-Test Split (80 : 20)

Pada tahap ini, dilakukan evaluasi performa metode *machine learning* dengan membagi *dataset* menjadi dua bagian, yaitu data untuk *training* sebesar 80% dari keseluruhan *dataset* dan data untuk *testing* sebesar 20% dari keseluruhan *dataset*. Selanjutnya, mengonversi setiap kata ke dalam bentuk vektor. Hal ini dilakukan karena *machine learning* tidak dapat menerima *input* dalam bentuk *string* tetapi hanya dapat menerima *input* dalam bentuk data numerik.

E. Pembentukan Metode

Pada tahap ini, dilakukan proses pembentukan metode dan tahap pembelajaran metode *hybrid* LSTM-SVM. Pada metode LSTM, dibangun sebuah model klasik dengan menerapkan beberapa *layer*, yaitu *embedding*, *spatial dropout*, LSTM, *dense* serta *activation*. Nilai pada masing-masing *layer* ditentukan dengan mencoba beberapa kemungkinan *hyperparameter* yang berbeda untuk menghasilkan nilai akurasi yang optimal dan diperoleh nilai yang digunakan, yaitu *spatial dropout* = 0,2; LSTM = 200; *dense* = 512 dan *activation* = *relu* dan *softmax*. Selanjutnya, membuat *class connector* dan melakukan inisiasi SVM ke dalam suatu *class* yang diberi nama ModelSVMWrapper serta penambahan *layer*. Kemudian, dilakukan inisiasi *input* LSTM dan *input* SVM.

F. Pengujian Metode

Pada tahap ini, pengujian metode dilakukan dengan mencari nilai akurasi berdasarkan nilai-nilai pada *layer* LSTM dan *class connector* yang telah didefinisikan ke dalam model sekuensial LSTM dengan *epochs* 100 dan *batch* 512.

Selanjutnya, merancang grafik akurasi dan grafik *loss* dari pengujian metode yang telah dilakukan untuk memberi gambaran mengenai akurasi maupun *loss* metode yang telah diuji. Gambar 9 menunjukkan bahwa *loss* yang dihasilkan mengalami kenaikan tidak lebih dari 20% yang mana grafik tersebut masih dapat dianggap sebagai hasil yang cukup baik dalam mempelajari model yang dikonstruksikan. Selanjutnya, dari Gambar 8 dan Gambar 9 dapat disimpulkan bahwa proses *training* pada kasus ini dikatakan cukup baik, tidak mengalami *overfitting* maupun *underfitting*.

G. Hasil Analisis Akurasi

Pada tahap ini, klasifikasi data dilakukan dengan mengkategorikan data menjadi empat jenis, yaitu *True Positive* (TP), *True Negative* (TN), *False Positive* (FP) dan *False Negative* (FN). *True Positive* (TP) adalah jumlah data

positif yang diklasifikasikan ke kelas positif. *True Negatif* (TN) adalah jumlah data negatif yang diklasifikasikan ke kelas negatif. *False Positive* (FP) adalah jumlah data positif yang diklasifikasikan ke kelas negatif. *False Negative* (FN) adalah jumlah data negatif yang diklasifikasikan ke kelas positif. Kemudian dilakukan tahap evaluasi akurasi, presisi, *recall* dan *f1-score*.

Akurasi mengukur ketepatan dan kemiripan hasil pada waktu yang sama dengan membandingkan terhadap nilai absolut. Nilai akurasi dapat dihitung dengan:

$$accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (11)$$

Presisi menunjukkan seberapa dekat perbedaan nilai pada saat dilakukan pengulangan pengukuran. Nilai presisi dapat dihitung dengan:

$$precision = \frac{TP}{TP+FP} \quad (12)$$

Recall merupakan tingkat keberhasilan sistem dalam menemukan kembali sebuah informasi. Nilai *recall* dapat dihitung dengan:

$$recall = \frac{TP}{TP+FN} \quad (13)$$

F1-Score dapat diinterpretasikan sebagai rata-rata dari presisi dan *recall*. Nilai *F1-Score* dapat dihitung dengan:

$$f1 - score = 2 \times \frac{precision \times recall}{precision + recall} \quad (14)$$

Dari penghitungan komputasi *confusion matrix* yang ditunjukkan oleh Gambar 10, diperoleh nilai akurasi dan nilai presisi dengan persentase yang sama besar, yaitu 94%. Kemudian untuk nilai *recall* dan nilai *f1-score* juga memiliki nilai persentase yang sama besar, yaitu 93% yang telah dirangkum pada Gambar 11.

Selaras dengan pembahasan performansi metode di atas, Tabel 7 merupakan tabel perbandingan hasil akurasi dari pengimplementasian tiga metode: LSTM, SVM, dan Hybrid LSTM-SVM. Pada Tabel 7 ditunjukkan bahwa nilai akurasi, presisi, *recall* dan *f1-score* dari metode Hybrid LSTM-SVM merupakan nilai yang paling tinggi dibandingkan dengan dua metode lainnya, yaitu LSTM dan SVM. Hal ini menunjukkan bahwa Metode Hybrid LSTM-SVM lebih baik dalam memprediksi berita online hoax COVID-19 di Indonesia dan dapat mengoptimalkan metode LSTM dan SVM.

H. Analisis Berita Online Hoax COVID-19 di Indonesia

Berita online merupakan informasi baru mengenai suatu hal yang sedang terjadi dan disajikan dalam bentuk cetak, siaran, internet bahkan dari mulut ke mulut kepada orang ketiga atau banyak orang. Seiring perkembangan zaman, terdapat berbagai media sosial dalam menyebarkan berita online, seperti *Weblog*, *Whatsapp*, *Instagram*, *Twitter* dan lain-lain, bahkan terdapat fitur dimana berita online dapat tersebar kepada banyak orang hanya dalam satu klik.

Dilihat dari kemudahan dalam mengakses media sosial, pengguna aktif media sosial terbiasa untuk berkomentar, memberi kritik dan membagikan informasi yang diperoleh.

Terlebih pada dua tahun terakhir ini, dimana munculnya virus baru yang disebut dengan *Coronavirus disease 2019* (COVID-19) yang kemudian menjadi perbincangan hangat setiap harinya. Beberapa dari kita, bahkan tidak

mengkonfirmasi atau mencari tahu terlebih dahulu mengenai kebenaran berita online yang tersebar. Hal ini dapat terjadi karena kepanikan yang muncul setelah membaca berita online tentang COVID-19 atau keinginan untuk menjadi orang pertama dalam menyebarkan berita online tentang COVID-19.

Hal semacam ini sangat tidak dibenarkan karena apabila berita online tentang COVID-19 yang tersebar merupakan berita online yang tidak terbukti benar, maka berita online tersebut bersifat bohong atau lebih dikenal dengan berita online hoax. Kemunculan berita online hoax tentang COVID-19 ini mampu mengakibatkan salah interpretasi dalam melawan pandemi COVID-19. Untuk menghindari hal tersebut, mengetahui terlebih dahulu kebenaran dari berita online tentang COVID-19 yang tersebar sangat diperlukan. Namun, dengan banyaknya informasi yang diberikan, ada kalanya menemui kesulitan dalam mencari informasi yang tepat untuk mengklarifikasi berita online tentang COVID-19 yang tersebar. Sehubungan dengan itu, dalam mengenali sebuah berita online merupakan berita online hoax atau berita online fakta dapat dilihat dari karakteristik kalimat berita online tentang COVID-19 tersebut.

Di bawah ini, diberikan beberapa contoh dari berita online hoax tentang COVID-19 di Indonesia:

1. Ketika rakyat muak dan jijik, tak ada lagi yang minta salaman dan selfie dengan Presiden Jokowi Widodo.
2. Corona Adalah Kedok dari Penyakit Radiasi.
3. Vaksin Sinovac Kadaluwarsa 25 Maret 2021 dan Dibuat Sebelum Pandemi.
4. China Lakukan Uji Coba Vaksin di Indonesia Karena Kekurangan Monyet.
5. Berita Terbaru Hari Ini 48 Orang Tewas Setelah Di Vaksin Corona, Agenda Jahat Rezim Terungkap!

Di bawah ini, diberikan beberapa contoh dari berita online fakta tentang COVID-19 di Indonesia:

1. Silaturahmi Virtual Upaya Melindungi Keluarga di Kampung Halaman Tanpa Mudik Lebaran.
2. Pasien Sembuh Terus Meningkat Mencapai 1.266.673 Orang.
3. BPOM Kawal Uji Klinik Vaksin COVID-19 di Indonesia.
4. Data Vaksinasi COVID-19 (Update per 14 April 2021).
5. Satgas Covid-19: Meskipun PSBB Transisi Harus 3M.

IV. KESIMPULAN DAN SARAN

Penelitian ini mengusulkan metode terkait *Hybrid Long Short Term Memory* dan *Support Vector Machine* pada data berita tentang COVID-19 di Indonesia untuk mendeteksi bahwa berita tersebut merupakan berita hoax atau berita fakta. LSTM digunakan untuk membangkitkan fitur yang representatif. Kemudian, SVM digunakan untuk mendeteksi berita hoax dengan memanfaatkan hasil representasi fitur yang telah dibangkitkan. Pada proses pengimplementasian Metode *Hybrid LSTM-SVM*, menerapkan beberapa *layer*. Nilai dari masing-masing *layer* ditentukan dengan mencoba beberapa kemungkinan *hyperparameter* yang berbeda untuk menghasilkan nilai akurasi dan nilai *loss* yang optimal. Hasilnya, nilai *loss* mengalami kenaikan namun tidak lebih dari 20% yang mana masih dapat dianggap sebagai hasil yang cukup baik dalam mempelajari model yang dikonstruksikan. Nilai akurasi mencapai 94%, yang mana hal ini menunjukkan

bahwa Metode *Hybrid LSTM-SVM* memiliki performa yang lebih tinggi dibandingkan dengan hanya mengimplementasikan Metode LSTM atau Metode SVM saja.

DAFTAR PUSTAKA

- [1] H. A. Rothan and S. N. Byrareddy, "The epidemiology and pathogenesis of coronavirus disease (COVID-19) outbreak," *J Autoimmun*, vol. 109, p. 102433, 2020, doi: 10.1016/J.JAUT.2020.102433.
- [2] A. Susilo *et al.*, "Coronavirus disease 2019: Tinjauan literatur terkini," *Jurnal Penyakit Dalam Indonesia*, vol. 7, no. 1, pp. 45–67, 2020, doi: 10.7454/JPDI.V7I1.415.
- [3] X. He, W. Hong, X. Pan, G. Lu, and X. Wei, "SARS-CoV-2 Omicron variant: Characteristics and prevention," *MedComm (Beijing)*, vol. 2, no. 4, pp. 838–845, 2021, doi: 10.1002/MCO2.110.
- [4] P. S. Reddy, D. E. Roy, P. Manoj, M. Keerthana, and P. v Tijare, "A study on fake news detection using Naïve Bayes, SVM, Neural Networks and LSTM," *Journal of Advanced Research in Dynamic and Control Systems*, vol. Volume 11, no. 06-Special Issue, pp. 942–947, 2019.
- [5] S. Sucipto, A. G. Tammam, and R. Indriati, "Hoax detection at social media with text mining clarification system-based," *JUPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)*, vol. 3, no. 2, 2018, doi: 10.29100/JUPI.V3I2.837.
- [6] P. Fulzele, R. Singh, N. Kaushik, and K. Pandey, "A hybrid model for music genre classification using LSTM and SVM," 2018. doi: 10.1109/IC3.2018.8530557.
- [7] G. Petmezas *et al.*, "Automated atrial fibrillation detection using a hybrid CNN-LSTM Network on imbalanced ECG datasets," *Biomed Signal Process Control*, vol. 63, p. 102194, 2021, doi: 10.1016/J.BSPC.2020.102194.
- [8] R. Fu, Z. Zhang, and L. Li, "Using LSTM and GRU neural network methods for traffic flow prediction," in *Proceedings - 2016 31st Youth Academic Annual Conference of Chinese Association of Automation, YAC 2016*, 2017, pp. 324–328. doi: 10.1109/YAC.2016.7804912.
- [9] S. Huang, C. A. I. Nianguang, P. Penzuti Pacheco, S. Narandes, Y. Wang, and X. U. Wayne, "Applications of Support Vector Machine (SVM) learning in cancer genomics," *Cancer Genomics Proteomics*, vol. 15, no. 1, p. 51, 2018, doi: 10.21873/CGP.20063.
- [10] E. Susilowati, M. K. Sabariah, and A. A. Gozali, "Impelentasi metode support vector machine untuk melakukan klasifikasi kemacetan lalu lintas pada twitter," *eProceedings of Engineering*, vol. 2, no. 1, pp. 1478–1484, 2015.
- [11] A. Cimino and F. Dell'Orletta, "Tandem LSTM-SVM approach for sentiment analysis," in *CEUR Workshop Proceedings*, 2016, vol. 1749, pp. 172–177. doi: 10.4000/BOOKS.AACCADEMIA.2003.