Volume 5, Nomor 1, Januari 2021, Page 24-33 ISSN 2614-5278 (media cetak), ISSN 2548-8368 (media online) Available Online at https://ejurnal.stmik-budidarma.ac.id/index.php/mib DOI 10.30865/mib.v5i1.2556



# Topic Modelling pada Sentimen Terhadap Headline Berita Online Berbahasa Indonesia Menggunakan LDA dan LSTM

#### Chairullah Naury, Dhomas Hatta Fudholi\*, Ahmad Fathan Hidayatullah

Fakultas Teknologi Industri, Magister Informatika, Universitas Islam Indonesia, Yogyakarta, Indonesia Email: <sup>1</sup>16917204@students.uii.ac.id, <sup>2</sup>hatta.fudholi@uii.ac.id, <sup>3</sup>fathan@uii.ac.id Email Penulis Korespondensi: hatta.fudholi@uii.ac.id

Abstrak— Media massa *online* adalah sumber informasi tercepat dan *up-to-date*. Sebuah model yang dapat memberikan pemetaan akan membantu dalam memilah informasi dengan lebih tepat. Pada penelitian ini, penulis menerapkan *topic modelling* terhadap hasil *sentiment analysis* pada *headline* berita *online* berbahasa Indonesia. Sumber data pada penelitian ini diperoleh dari media massa *online* berbahasa Indonesia. Data yang terkumpul dianalisis sentimennya dengan menggunakan metode *Long Short-term Memory* (LSTM), sehingga diperoleh tajuk-tajuk berita dengan sentimen positif, negatif, dan netral. Klasifikasi yang diperoleh dari hasil proses sentimen analisis tersebut dilanjutkan dengan proses pemodelan topik menggunakan metode *Latent Dirichlet Allocation* (LDA) dan divisualisasikan dalam bentuk *wordcloud* dan *intertopic distance map* (pyLDAVis) untuk mengetahui keterkaitan satu topik dengan topik lainnya. Proses analisis sentimen menghasilkan model dengan tingkat akurasi 71.13% dan proses pemodelan topik berhasil menyajikan topik-topik yang mudah dinterpretasikan sehingga dapat diketahui kesimpulan tentang suatu isu.

Kata Kunci: Analisis Sentimen; Long Short-Term Memory; Headline Berita Online; Pemodelan Topik, Latent Dirichlet Allocation.

**Abstract**—The online mass media is the source of the fastest and up-to-date information. A model that can provide mapping will help in sorting out information more precisely. In this study, the authors applied topic modeling to the results of sentiment analysis on online news headlines in Indonesian. Sources of data in this study were obtained from online mass media in Indonesian. The data collected were analyzed for sentiment using the Long Short-term Memory (LSTM) method, in order to obtain news headlines with positive, negative, and neutral sentiments. The classification obtained from the results of the sentiment analysis process is continued with the topic modeling process using the Latent Dirichlet Allocation (LDA) method and visualized in the form of wordcloud and intertopic distance map (pyLDAVis) to determine the relationship between one topic and another. The result of sentiment analysis is a model with 71.13% of accuracy level and the results of topic modeling are in the form of some topics that are easy to interpret.

Keywords: Sentiment Analysis; Long Short-Term Memory; Online News Headline; Topic Modelling; Latent Dirichlet Allocation

# 1. PENDAHULUAN

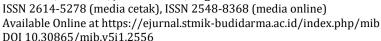
Headline atau tajuk berita adalah kalimat pendek atau frasa yang ditempatkan secara mencolok pada sebuah berita dengan menggunakan huruf yang menonjol [1]. Headline news atau berita utama adalah berita yang disepakati oleh dewan redaksi dan paling layak untuk ditampilkan di halaman depan surat kabar dan menggunakan judul yang dicetak dengan huruf lebih besar dari berita lainnya dan menarik mengundang rasa penasaran para pembaca [2]. Pesan pertama yang dibaca oleh pembaca berita adalah judul berita sebelum membaca keseluruhan informasi yang disajikan dalam sebuah berita. Ringkasan informasi pada sebuah berita pada umumnya terangkum di bagian judul berita. Struktur kalimat dalam judul berita dibuat efisien dengan penekanan pada unsur siapa dan apa yang dilakukan. Kata kerja, kata benda, kata sifat, dan kata keterangan dipilih yang singkat dan tajam untuk menyampaikan pesan kepada pembaca [3].

Penelitian ini membahas tentang topic modelling pada hasil sentiment analysis terhadap headline atau tajuk berita yang dipublikasikan di media berita online berbahasa Indonesia. Topic modelling adalah salah satu teknik di dalam Natural Language Processing (NLP) untuk menganalisis teks [4] yang berupa algoritma untuk mengidentifikasi pola-pola tersembunyi dari serangkaian kata dengan menggunakan teknik pendistribusian katakata di dalam sekumpulan dokumen. Output dari topic modelling berupa sekumpulan topik yang terdiri dari beberapa gugus kata yang muncul bersamaan di dalam dokumen berdasarkan pola tertentu [5]. Metode topic modelling banyak ragamnya, yaitu Latent Semantic Allocation (LSA) [6], Probabilistic Latent Semantic Analysis (PLSA) [7], dan Latent Dirichlet Allocation (LDA) [8]. Setiap model memiliki beberapa tahapan, dan setiap tahapan akan berpengaruh terhadap hasil yang diharapkan [9]. Penulis di dalam penelitian ini menggunakan metode Latent Dirichlet Allocation (LDA) untuk pemrosesan pemodelan topik terhadap hasil analisis sentimen tajuk berita online berbahasa Indonesia. Sentiment analysis juga merupakan salah satu teknik di dalam NLP. Sentiment analysis dapat dilakukan dengan menggunakan dua pendekatan, yaitu Machine Learning Approach dan Lexicon-based Approach [10]. Pendekatan yang digunakan pada penelitian ini adalah pendekatan Machine Learning, yaitu menggunakan metode Recurrent Neural Network.

Recurrent Neural Network (RNN) adalah salah satu dari metode Neural Network yang digunakan untuk menyelesaikan persoalan terkait dengan NLP. Metode RNN menggunakan perhitungan bobot secara berulang, sehingga nilai akurasi yang diperoleh lebih lebih baik dari jaringan saraf tiruan sederhana [11]. Metode RNN dipakai dengan tujuan agar mesin mampu memahami bahasa manusia termasuk cara berkomunikasi,

Chairullah Naury, Copyright © 2021, MIB, Page 24 Submitted: 28/10/2020; Accepted: 25/11/2020; Published: 22/01/2021

Volume 5, Nomor 1, Januari 2021, Page 24-33





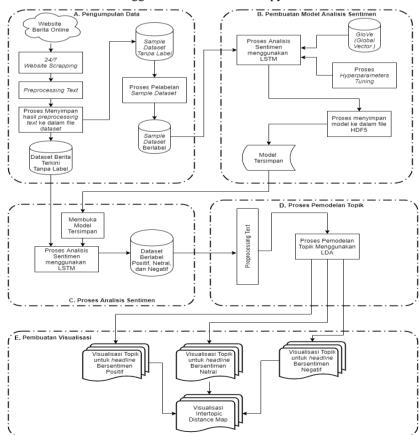
mendengarkan, mengenali percakapan, dan memahami tata bahasa tertentu [12]. Jenis RNN yang digunakan di dalam penelitian ini adalah *Long Short Term Memory* (LSTM). LSTM terbukti dapat menutupi kekurangan dari RNN terkait ketidakmampuannya di dalam menyimpan memori agar dapat diseleksi dan dan menambah mekanisme *attention* agar setiap kata sesuai dengan konteksnya [11].

Berawal dari posisi berita yang merupakan salah satu bentuk informasi yang banyak dicari oleh masyarakat [13] dan mulai dipublikasikan secara *online* dalam bentuk *website* atau portal berita [14], penulis mengangkat domain ini ke dalam penelitian yang dilakukan. Secara umum media massa memiliki banyak fungsi, yaitu: (1) penyebaran informasi, (2) penyampaian interpretasi atau opini terhadap suatu informasi, (3) pembentukan kesepakatan, (4) korelasi/penghubung antar kelompok masyarakat, (5) transmisi warisan budaya, (6) mengekspresikan nilai-nilai budaya untuk melestarikan identitas masyarakat [15]. Terlebih lagi, penulis melakukan *literature review* terhadap beberapa artikel ilmiah yang berkaitan dengan pemodelan topik dan analisis sentimen untuk menunjukkan manfaat dari impelementasi kedua metode tersebut. Pemodelan topik menggunakan metode LDA untuk mengetahui persepsi masyarakat terhadap kebijakan yang digulirkan oleh pembuat kebijakan diteliti di [16]. Pada penelitian lain, bagaimana melakukan prediksi harga saham di pasar modal dengan memanfaatkan data harga saham dan sentimen berita di media massa menggunakan metode LSTM [17] juga telah dilakukan dan berhasil memberikan hasil yang memuaskan.

Berpijak pada hasil *literature review* dan gap dari penelitian sebelumnya, penulis melakukan penelitian yang bertujuan untuk menghasilkan model yang dapat menganalisis topik berdasarkan hasil sentimen analisis pada *headline* berita *online* berbahasa Indonesia. Topik-topik yang dihasilkan akan dikelompokkan berdasarkan topik-topik berita bersentimen positif, netral, dan negatif. Kontribusi penelitian ini adalah memperkaya literatur penelitian di bidang *topic modelling based sentiment analysis*. Kontribusi lainnya adalah agar masyarakat dan pemangku kebijakan dapat memperoleh *insight* bermanfaat dari topik-topik yang dibicarakan di media massa terkait dengan suatu isu berita. *Insight* yang diperoleh diharapkan dapat menjadi masukan bagi masyarakat maupun para pemangku kebijakan di dalam pembuatan kebijakan terkait penanganan suatu isu berita.

### 2. METODOLOGI PENELITIAN

Tahapan-tahapan yang dilaksanakan pada penelitian ini ditunjukkan pada Gambar 1. Tahap pertama adalah pengumpulan data, tahap kedua pembuatan model sentimen analisis, tahap ketiga proses analisis sentimen menggunakan metode LSTM, tahap keempat proses pemodelan topik menggunakan metode LDA, dan tahap kelima adalah pembuatan visualisasi menggunakan *wordcloud* dan pyLDAVis.



Gambar 1. Diagram Alur Penelitian

Chairullah Naury, Copyright ©2021, MIB, Page 25 Submitted: 28/10/2020; Accepted: 25/11/2020; Published: 22/01/2021

**Volume 5, Nomor 1, Januari 2021, Page 24-33**ISSN 2614-5278 (media cetak), ISSN 2548-8368 (media online)
Available Online at https://ejurnal.stmik-budidarma.ac.id/index.php/mib
DOI 10.30865/mib.v5i1.2556



#### 2.1 Pengumpulan Data

Penelitian ini memerlukan data tajuk berita dari media massa *online*. Untuk memperoleh data yang berupa tajuk berita *online* berbahasa Indonesia, penulis menggunakan cara *scrapping* pada situs-situs berita *online*. Situs berita *online* berita berbahasa Indonesia yang penulis gunakan sebagai sumber data adalah antaranews.com, cnnindonesia.com, detik.com, jpnn.com, kompas.com, liputan6.com, merdeka.com, okezone.com, rmol.id, sindonews.com, suara.com, dan tempo.co. Situs-situs berita online tersebut dipilih karena situs tersebut dikelola oleh perusahaan yang benar-benar bergerak di bidang jurnalistik.

Studi kasus pada penelitian ini menggunakan data hasil dari query data tajuk berita yang mengandung term "coronavirus disease 2019" atau "Covid-19". Coronavirus disease 2019 adalah penyakit yang disebabkan virus corona yang pertama kali menjangkiti warga Wuhan di Tiongkok pada akhir tahun 2019 yang lalu dan menyebabkan pandemi (wabah mendunia) serta menjadi masalah utama kesehatan masyarakat dunia [18]. Coronavirus ini dapat menginfeksi masyarakat dari berbagai tingkatan usia. Kelompok masyarakat yang paling beresiko terhadap dampak infeksi Covid-19 ini adalah kelompok usia tua (usia 60 tahun keatas) dan kelompok masyarakat yang memiliki riwayat penyakit kronis seperti diabetes, gangguan jantung, paru-paru, dan kanker [19].

Data yang diperoleh dari hasil proses *scrapping* situs berita *online* penulis bagi menjadi dua bagian. Dataset bagian pertama penulis gunakan sebagai *sample dataset* atau data latih untuk pembuatan model *sentiment analysis*, dan bagian kedua adalah data yang selalu mengalir dan terus menerus tersimpan ke dalam dataset yang berisi data *headline* berita terkini. Dataset bagian kedua inilah yang nantinya akan digunakan di dalam proses analisis sentimen, pemodelan topik dan visualisasi. Dataset bagian pertama hasil dari *scrapping* situs berita *online* masih berupa *unlabeled dataset*, yaitu dataset yang belum memiliki label positif, negatif, dan netral. Proses pelabelan *sample dataset* ini dilakukan secara manual sehingga hasil akhirnya berupa *sample dataset* berlabel positif, negatif, dan netral.

#### 2.2 Pembuatan Model Analisis Sentimen

Tahapan kedua yang dilakukan adalah pembuatan model analisis sentimen. Proses pembuatan model pada penelitian ini memerlukan *pre-trained global vector for word representation* berbahasa Indonesia yang penulis peroleh dari *fasttext* yang memuat 2.000.000 *term* dengan dimensi sebanyak 50 dimensi. Pembuatan model dilakukan dengan melatih *dataset* melalui beberapa skenario *hyperparameters tuning*. *Hyperparameters tuning* adalah proses pemilihan dan pengoptimalan berbagai *hyperparameter* di dalam jaringan syaraf tiruan, termasuk LSTM. Pemilihan kombinasi beberapa *hyperparameter* dapat menghasilkan model dengan performa yang berbeda-beda. Proses pembuatan model ini dilakukan dengan menyesuaikan berbagai parameter seperti jumlah *layer, recurrent units, input size, batch size, epoch*, maupun *dropout* [20].

Tabel 1 menunjukkan skenario pelatihan dataset. Terdapat 14 skenario eksperimen yang memadupadankan berbagai *hyperparameter* seperti jumlah lapisan LSTM, jumlah *recurrent units*, jumlah *dropout*, banyaknya *batch size*, dan maksimal *epoch*. Hasil pelatihan *dataset* yang dilakukan dalam beberapa eksperimen tersebut selanjutnya dievaluasi untuk diketahui tingkat akurasi modelnya. Model dengan tingkat akurasi yang tinggi disimpan sebagai model yang akan digunakan untuk proses analisis sentimen berikutnya.

Eksperimen Jumlah Recurrent **Dropouts** Max Batch **LSTM** Units Size **Epoch** Layer #1 128 16 64 1 #2 1 128 64 32 #3 1 128 128 16 #4 1 128 128 32 #5 2 32;16 0.2 64 16 2 64 #6 32;16 0.2 32 2 128 #7 32;16 0.2 16 2 #8 32;16 0.2 128 32 2 #9 64:16 0.2 64 16 2 #10 256:16 0.2 64 16 #11 3 32:16:8 0.5 64 16 #12 3 32;16;8 0.5 64 32 3 #13 32;16;8 128 16 0.5 3 #14 32:16:8 0.5 128 32

**Tabel 1.** Skenario Hyperparameters Tuning

#### 2.3 Proses Analisis Sentimen

Pada tahapan ini, penulis melakukan proses analisis sentimen terhadap *unlabeled dataset*. *Dataset* tajuk berita ini belum memiliki label positif, negatif, maupun netral. *Unlabeled dataset* diperoleh dari hasil *scrapping* situs-situs berita *online*. Proses analisis sentimen pada tahap ini menggunakan model yang sudah tersimpan sebelumnya,

Chairullah Naury, Copyright © 2021, MIB, Page 26 Submitted: 28/10/2020; Accepted: 25/11/2020; Published: 22/01/2021

Volume 5, Nomor 1, Januari 2021, Page 24-33

ISSN 2614-5278 (media cetak), ISSN 2548-8368 (media online) Available Online at https://ejurnal.stmik-budidarma.ac.id/index.php/mib DOI 10.30865/mib.v5i1.2556



sehingga diperoleh dataset baru yang sudah memiliki label positif, negatif, dan netral. Dataset yang sudah berlabel ini selanjutnya akan dipakai untuk proses pemodelan topik.

### 2.4 Proses Pemodelan Topik

Pemodelan Topik dilakukan dengan menggunakan metode Latent Dirichlet Allocation (LDA). Proses pemodelan topik dilakukan untuk masing-masing dataset berlabel positif, dataset berlabel negatif, dan dataset berlabel netral. Sehingga diperoleh topik-topik apa saja yang dibicarakan media massa *online* terkait suatu isu berita, baik itu topik vang berasal dari tajuk berita dengan sentimen positif, negatif, dan netral.

#### 2.5 Visualisasi

Topik-topik yang diperoleh dari hasil pemodelan topik selanjutnya divisualisasikan dalam bentuk wordcloud [21] agar lebih mudah untuk melihat insight topik pembicaraan media massa online. Selain wordcloud, visualisasi lainnya adalah diagram *intertopic distance map* [22] untuk melihat jarak dan irisan keterkaitan antar topik.

#### 3. HASIL DAN PEMBAHASAN

Bagian ini menjelaskan hasil dari penelitian yang dilakukan. Hasil yang didapatkan dielaborasikan sesuai dengan tahapan penelitian yang dilakukan.

### 3.1 Hasil Pengumpulan Data

Untuk pengumpulan data, penulis menggunakan teknik scrapping situs-situs berita online. Penulis menggunakan kode program dalam bahasa python untuk mendapatkan tajuk berita dengan menggunakan library BeautifulSoup. Library tersebut mempermudah proses parsing elemen-elemen html seperti tag div, h1, dan h2 yang umumnya digunakan sebagai tag judul berita pada sebuah website.

Data tajuk berita yang berhasil di-scrapping selanjutnya disimpan ke dalam database mysql. Database mysql dipilih untuk mempermudah query pada saat pencarian keyword tertentu untuk dianalisis sentimennya. Semua data tajuk berita yang tersimpan di database belum memiliki label sentimen positif, negatif, maupun netral. Untuk keperluan membuat sample dataset, maka dipilih 2500 record secara acak dan disimpan ke dalam tabel unlabeled sample dataset. Ouery SOL yang digunakan untuk memilih 2500 record secara acak sebagai berikut: INSERT INTO "unlabeled sample dataset" SELECT \* FROM 'headlines' ORDER BY RAND() LIMIT 2500; Hasil dari query SQL tersebut adalah sebuah tabel bernama "unlabeled\_sample\_dataset" yang berisi 2500 record headline berita yang diambil secara acak dari tabel "headlines".

Proses selanjutnya adalah melakukan pelabelan dataset secara manual. Tabel "Unlabeled\_sample\_dataset" yang belum memiliki label diekspor ke dalam format file CSV dan dibuka menggunakan software spreadsheet, kemudian ditentukan label dari masing-masing tajuk berita, sehingga setiap headline berita akhirnya memiliki label negatif, netral, atau positif. Nama file "Unlabeled\_sample\_dataset.csv" selanjutnya di-rename menjadi "sample dataset.csv" dan file ini akan menjadi dataset untuk proses pembuatan model analisis sentimen.

#### 3.2 Hasil Pembuatan Model Analisis Sentimen

Sample dataset yang diperoleh dari proses pengumpulan data kemudian melalui proses preprocessing text sebelum digunakan di dalam pembuatan model. Langkah-langkah Preprocessing text yang dilakukan adalah 1) membuang stopwords; 2) membuang karakter non-ascii, url, kata sambung, dan angka-angka; 3) membuang kalimat yang hanya terdiri dari satu kata.

Penulis melakukan 14 eksperimen untuk pembuatan model analisis sentimen. Setiap eksperimen bertujuan untuk menguji tingkat akurasi model dengan menyesuaikan beberapa parameter model, seperti jumlah *layer*, recurrent units tiap layer, dropout tiap layer, batch size, dan epoch. Eksperimen dilakukan dengan menjalani langkah-langkah utama berikut: pembuatan model LSTM, melakukan training pada model, dan mencari tingkat akurasi.

Layer (type)	Output Shape	Param #
embedding (Embedding)	(None, 7, 50)	215600
lstm (LSTM)	(None, 7, 256)	314368
lstm_1 (LSTM)	(None, 16)	17472
dense (Dense)	(None, 3)	51
Total params: 547,491 Trainable params: 331,891 Non-trainable params: 215,6	500	

Gambar 2. Arsitektur model LSTM

Arsitektur model LSTM dibangun pertama kali dengan susunan seperti pada Gambar 2. Setelah pembuatan model LSTM, langkah berikutnya adalah melakukan training dataset pada model LSTM yang sudah didefinisikan sebelumnya. Proses training dataset ini dilakukan dalam beberapa epoch sesuai skenario hyperparameters tuning.

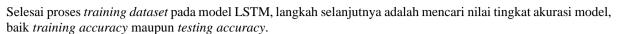
> Chairullah Naury, Copyright ©2021, MIB, Page 27 Submitted: 28/10/2020; Accepted: 25/11/2020; Published: 22/01/2021

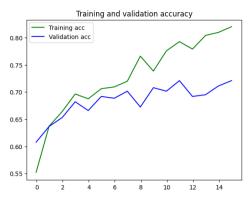
Volume 5, Nomor 1, Januari 2021, Page 24-33

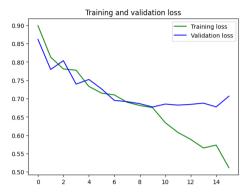
ISSN 2614-5278 (media cetak), ISSN 2548-8368 (media online)

Available Online at https://ejurnal.stmik-budidarma.ac.id/index.php/mib

DOI 10.30865/mib.v5i1.2556







Gambar 3. Grafik Training and Validation Accuracy Gambar 4. Grafik Training and Validation Loss

Gambar 3 dan gambar 4 menunjukkan hasil dari proses training dataset. Hasil dari beberapa eksperimen sesuai skenario hyperparameters tuning terhadap model LSTM ditunjukkan pada Tabel 2. Dengan memperhatikan hasil test accuracy, penulis memilih model eksperimen #10 sebagai model yang akan digunakan untuk proses analisis sentimen.

**Tabel 2.** Hasil Hyperparameters Tuning

Eksp#	#1	#2	#3	#4	#5	#6	#7	#8	#9	#10	#11	#12	#13	#14
Acc (%)	66.35	67.30	67.69	69.41	68.64	68.64	66.35	67.30	70.36	71.13	64.82	65.97	63.48	65.77

#### 3.3 Hasil Analisis Sentimen

Proses analisis sentimen pada tahap ini menggunakan model yang tersimpan sebelumnya dan dataset tajuk berita terkini yang belum memiliki label positif, negatif, dan netral. Proses analisis sentimen ini ditunjukkan pada gambar 5. Hasil proses analisis sentimen tajuk berita *online* disimpan ke database menjadi *dataset* untuk proses pemodelan topik dengan menggunakan metode Latent Dirichlet Allocation.

```
Sedang memroses data ke-529 dari 27103 baris data
['Tangkal Virus Corona, RI Akan Tutup Penerbangan ke Seluruh Cina']
positive
Sedang memroses data ke-530 dari 27103 baris data
['Wabah Virus Corona, Lion Air Terbang dari Wuhan ke Denpasar']
```

Gambar 5. Proses analisis sentimen tajuk berita online

#### 3.4 Hasil Pemodelan Topik

Penulis melakukan pemodelan topik terhadap dataset hasil analisis sentimen pada proses sebelumnya. Penulis menggunakan rentang tanggal 05 April 2020 hingga tanggal 12 April 2020 dari dataset untuk dibuat pemodelan topiknya. Hasil ekstraksi topik direpresentasikan pada Tabel 3 untuk pemodelan topik terhadap tajuk berita yang bersentimen positif, Tabel 4 untuk pemodelan topik terhadap tajuk berita bersentimen negatif, dan Tabel 5 untuk pemodelan topik terhadap tajuk berita bersentimen netral.

Tabel 3. Hasil Ekstraksi Topik Tajuk Berita Bersentimen Positif

Topic ID	Word Distribution
Topic: 0	virus, sembuh, pasien_positif, dinyatakan_sembuh, tenaga_medis, meninggal, pasien,
	sembuh_meninggal, rapid_test, bertambah.
Topic: 1	update, rumah_sakit, jakarta, meninggal, sembuh, tes, lawan, warga, indonesia, cegah.
Topic: 2	pandemi, orang, pasien_positif, bertambah, virus, wabah, pasien, sembuh, indonesia, sumut.
Topic: 3	dunia, pasien, darurat, juta, perangi, bantu, sembuh, bertambah, tenaga_medis, total.
Topic: 4	pasien_positif, virus, indonesia, pasien, april, wabah, rapid_test, orang, ahli, cegah.
Topic: 5	pdp, meninggal, imbas, update, pasien_positif, karantina, rs, kota, dunia, pasien.

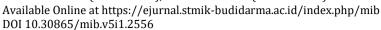
Tabel 4. Hasil Ekstraksi Topik Tajuk Berita Bersentimen Negatif

Topic ID	Word Distribution
Topic: 0	virus, meninggal_dunia, kematian, terdampak, masyarakat, dki, darurat, psbb, meninggal, akibat.
Topic: 1	virus, wabah, pemerintah, warga, akibat, indonesia, ribu, polisi, pasien, jokowi.
Topic: 2	orang, korban, meninggal, sembuh, jenazah, pasien, as, virus, china, wabah.

Chairullah Naury, Copyright ©2021, MIB, Page 28 Submitted: 28/10/2020; Accepted: 25/11/2020; Published: 22/01/2021

Volume 5, Nomor 1, Januari 2021, Page 24-33

ISSN 2614-5278 (media cetak), ISSN 2548-8368 (media online)





Topic ID	Word Distribution
Topic: 3	lawan, pandemi, orang, imbas, pemerintah, virus, tutup, rs, lockdown, perusahaan.
Topic: 4	pandemi, data, virus, pm_inggris, pasien, warga, dunia, karyawan, akibat, pemda.
Topic: 5	positif, tenaga_medis, virus, dokter, pemerintah, cegah, warga, penanganan, penyebaran, malaysia

Tabel 5. Hasil Ekstraksi Topik Tajuk Berita Bersentimen Netral

Topic ID	Word Distribution
Topic: 0	pasien, as, cegah_penyebaran, dampak, virus, akibat, video, jakarta, menag, orang
Topic: 1	wabah, virus, jokowi, rumah_sakit, imbas, tangani, sembuh, sembuh_meninggal, anggaran, china.
Topic: 2	pandemi, meninggal, sembuh, indonesia, orang, pasien, virus, april, update, positif.
Topic: 3	pandemi, positif, kota, virus, sembuh, rumah, ribu, daftar, indonesia, april.
Topic: 4	positif, gugus_tugas, pasien, bantuan, warga, terdampak, virus, online, sembuh, menteri.
Topic: 5	pemerintah, pandemi, meninggal_dunia, tenaga_medis, pasien, anak, dunia, cerita, nasib, darurat.

#### 3.5 Hasil Visualisasi

Hasil ekstraksi topik-topik dari proses pemodelan topik selanjutnya akan divisualisasikan dalam bentuk wordcloud dan intertopic distance map. Visualisasi Wordcloud digunakan untuk mempermudah memperoleh insight dari topik-topik yang dihasilkan, dan intertopic distance map digunakan untuk mengetahui jarak antar topik dan keterkaitan satu topik dengan topik lainnya. Insight yang diperoleh dari hasil visualisasi ditunjukkan pada tabel 6.

### 3.5.1 Visualisasi topik pada hasil analisis sentimen positif

Visualisasi pemodelan topik pada hasil analisis sentimen tajuk berita berlabel positif untuk rentang tanggal 05 April 2020 hingga 12 April 2020 berupa *wordcloud* dan *intertopic distance map*. Visualisasi *wordcloud* untuk topik tajuk berita bersentimen positif disajikan pada gambar 6, gambar 7, gambar 8, gambar 9, gambar 10, dan gambar 11. Sedangkan visualisasi *intertopic distance map* disajikan pada gambar 12.



**Gambar 6.** Wordcloud topik #0



Gambar 7. Wordcloud topik #1



Gambar 8. Wordcloud topik #2



Gambar 9. Wordcloud topik #3



Gambar 10. Wordcloud topik #4

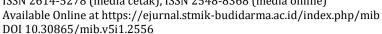


Gambar 11. Wordcloud topik #5

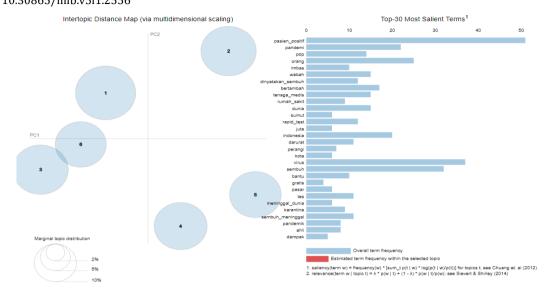
Chairullah Naury, Copyright © 2021, MIB, Page 29 Submitted: 28/10/2020; Accepted: 25/11/2020; Published: 22/01/2021

Volume 5, Nomor 1, Januari 2021, Page 24-33

ISSN 2614-5278 (media cetak), ISSN 2548-8368 (media online)







Gambar 12. Intertopic Distance Map

#### 3.5.2 Visualisasi Topik Pada Hasil Analisis Sentimen Negatif

Visualisasi pemodelan topik pada hasil analisis sentimen tajuk berita berlabel negatif untuk rentang tanggal 05 April 2020 hingga 12 April 2020 berupa *wordcloud* dan *intertopic distance map*. Visualisasi *wordcloud* untuk topik tajuk berita bersentimen negatif disajikan pada gambar 13, gambar 14, gambar 15, gambar 16, gambar 17, dan gambar 18. Sedangkan visualisasi *intertopic distance map* disajikan pada gambar 19.



Gambar 13. Wordcloud topik #0



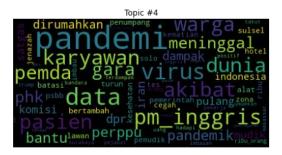
Gambar 14. Wordcloud topik #1



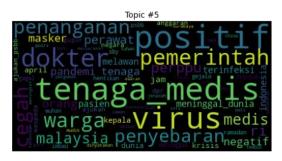
Gambar 15. Wordcloud topik #2



Gambar 16. Wordcloud topik #3



**Gambar 17.** *Wordcloud* topik #4



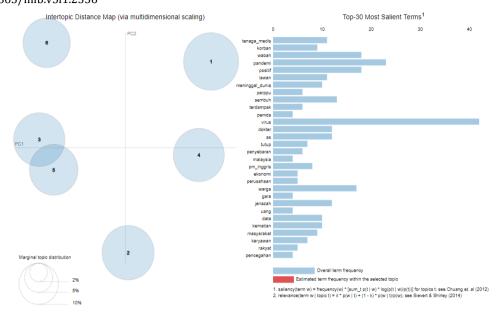
Gambar 18. Wordcloud topik #5

Volume 5, Nomor 1, Januari 2021, Page 24-33

ISSN 2614-5278 (media cetak), ISSN 2548-8368 (media online)

Available Online at https://ejurnal.stmik-budidarma.ac.id/index.php/mib DOI 10.30865/mib.v5i1.2556





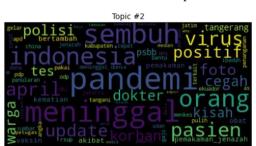
Gambar 19. Intertopic Distance Map

### 3.5.3 Visualisasi topik pada hasil analisis sentimen netral

Visualisasi pemodelan topik pada hasil analisis sentimen tajuk berita berlabel netral untuk rentang tanggal 05 April 2020 hingga 12 April 2020 berupa *wordcloud* dan *intertopic distance map*. Visualisasi *wordcloud* untuk topik tajuk berita bersentimen netral disajikan pada gambar 20, gambar 21, gambar 22, gambar 23, gambar 24, dan gambar 25. Sedangkan visualisasi *intertopic distance map* disajikan pada gambar 26.



Gambar 20. Wordcloud topik #0



Gambar 22. Wordcloud topik #2



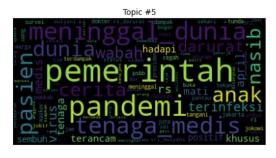
Gambar 24. Wordcloud topik #4



Gambar 21. Wordcloud topik #1



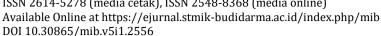
Gambar 23. Wordcloud topik #3



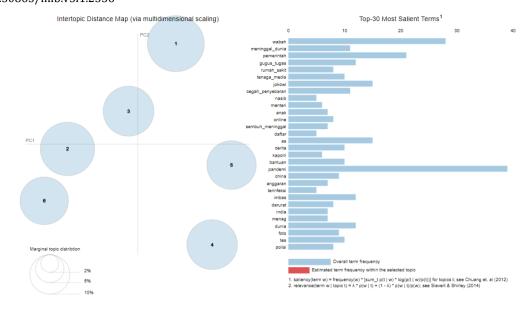
Gambar 25. Wordcloud topik #5

Volume 5, Nomor 1, Januari 2021, Page 24-33

ISSN 2614-5278 (media cetak), ISSN 2548-8368 (media online)







Gambar 26. Intertopic Distance Map

Insight yang dapat diperoleh dari visualisasi wordcloud topik headline bersentimen positif, negatif, dan netral ditunjukkan pada Tabel 6.

Insight Sentimen Positif Insight Sentimen Negatif Insight Sentimen Netral Topic ID Topic 0 Berita kesembuhan pasien covid Berita tentang kematian dampak Berita pencegahan penyebaran dari virus virus. Topic 1 Berita update data pasien sembuh Berita tentang penanganan Berita penanganan wabah virus. dan meninggal masalah pekerja yang kena PHK dampak wabah virus. Topic 2 Berita dampak pandemi dalam Berita update data pasien Berita pasien meninggal akibat lingkup Indonesia. meninggal dan sembuh bulan April. virus. Topic 3 Berita kondisi darurat dalam Berita tentang lockdown imbas Berita tentang kota-kota di lingkup dunia. pandemic. Indonesia yang terkena pandemi virus. Topic 4 Berita rapid test dan pencegahan Berita tentang karyawan Berita upaya gugus tugas di dalam menyebarnya virus di Indonesia. dirumahkan akibat pandemi. mencegah dampak virus. Topic 5 Berita pasien positif, PDP, dan Berita pemerintah hadapi pandemi Berita tentang pemerintah, meninggal di beberapa kota di dokter, dan tenaga medis cegah dunia. penyebaran virus.

Tabel 6. Insight Hasil Ekstraksi Topik

# 4. KESIMPULAN

Analisis sentimen yang dilaksanakan pada penelitian ini menggunakan metode Long Short-term Memory (LSTM) dengan memanfaatkan pre-trained global vector for word representation berbahasa Indonesia yang diperoleh dari fasttext menghasilkan model dengan nilai akurasi tertinggi sebesar 71.13% dan nilai akurasi terendah sebesar 63.48%. Pada setiap eksperimen hyperparameters tuning, proses ini menghasilkan nilai akurasi yang berbedabeda sesuai dengan parameter-parameter yang dipasang, yaitu jumlah lapisan LSTM, jumlah recurrent units. dropouts, ukuran batch, dan maksimum epoch. Pemodelan topik pada penelitian ini menggunakan metode Latent Dirichlet Allocation (LDA) dapat menghasilkan topik-topik yang dibicarakan oleh media online, baik itu topiktopik yang bersentimen positif, negatif, maupun netral. Topik-topik yang diperoleh dari pemodelan topik terhadap hasil analisis sentimen pada tajuk berita online berbahasa Indonesia ini dapat diinterpretasikan dengan baik. Dari hasil visualisasi yang disajikan dapat diperoleh topik-topik yang memiliki insight dari suatu isu berita.

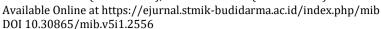
### REFERENCES

- Badan Pengembangan Bahasa dan Perbukuan, "Kamus Besar Bahasa Indonesia," Kementerian Pendidikan dan Kebudayaan, 2016. https://kbbi.kemdikbud.go.id/.
- M. Wandik, A. M.Golung, and H.Mulyono, "Proses Penentuan Headline Surat Kabar (Studi Pada Surat Kabar Harian Manado Post)," E-Journal Acta Diurna, vol. VI, no. 2, 2017.
- M. Winiharti, "Analisis Diksi Pada Judul Berita Utama Surat Kabar yang Memberitakan Rapat Pansus DPR RI untuk Kasus Bank Century," J. Penelit. Hum., vol. 12, no. 1, pp. 19-31, 2011.

Chairullah Naury, Copyright ©2021, MIB, Page 32 Submitted: 28/10/2020; Accepted: 25/11/2020; Published: 22/01/2021

Volume 5, Nomor 1, Januari 2021, Page 24-33

ISSN 2614-5278 (media cetak), ISSN 2548-8368 (media online)





- [5] C. Jacobi, W. Van Atteveldt, and K. Welbers, "Quantitative analysis of large amounts of journalistic texts using topic modelling," *Digit. Journal.*, vol. 4, no. 1, pp. 89–106, 2016, doi: 10.1080/21670811.2015.1093271.
- [6] S. Bergamaschi and L. Po, "Comparing LDA and LSA topic models for content-based movie recommendation systems," Lect. Notes Bus. Inf. Process., vol. 226, pp. 247–263, 2015, doi: 10.1007/978-3-319-27030-2\_16.
- [7] Y. Lu, Q. Mei, and C. X. Zhai, "Investigating task performance of probabilistic topic models: An empirical study of PLSA and LDA," *Inf. Retr. Boston.*, vol. 14, no. 2, pp. 178–203, 2011, doi: 10.1007/s10791-010-9141-9.
- [8] D. M. Blei, B. B. Edu, A. Y. Ng, A. S. Edu, M. I. Jordan, and J. B. Edu, "Latent Dirichlet Allocation," CrossRef List. Deleted DOIs, vol. 1, pp. 993–1022, 2003, doi: 10.1162/jmlr.2003.3.4-5.993.
- [9] L. Hagen, "Content analysis of e-petitions with topic modeling: How to train and evaluate LDA models?," *Inf. Process. Manag.*, vol. 54, no. 6, pp. 1292–1307, 2018, doi: 10.1016/j.ipm.2018.05.006.
- [10] W. Medhat, A. Hassan, and H. Korashy, "Sentiment analysis algorithms and applications: A survey," *Ain Shams Eng. J.*, vol. 5, no. 4, pp. 1093–1113, 2014, doi: 10.1016/j.asej.2014.04.011.
- [11] K. Ivanedra and M. Mustikasari, "Implementasi Metode Recurrent Neural Network Pada Text Summarization Dengan Teknik Abstraktif the Implementation of Text Summarization With Abstractive Techniques Using Recurrent Neural Network Method," vol. 6, no. 4, pp. 377–382, 2019, doi: 10.25126/jtiik.201961067.
- [12] R. Primartha, Belajar Machine Learning Teori dan Praktik. Bandung: Penerbit INFORMATIKA, 2018.
- [13] B. K. Wangsa, D. Utomo, and S. Nugroho, "Sistem Peringkas Berita Otomatis berbasis Text Mining menggunakan Generalized Vector Space Model: Studi Kasus Berita diambil dari Media Massa Online," *Techne J. Ilm. Elektrotek.*, vol. 13, pp. 231–241, 2014.
- [14] M. R. Pratiwi, "Peran ICT bagi Organisasi Media Massa dan Budaya Masyarakat," *Komunikator*, vol. 6, no. 5, p. 21, 2014, [Online]. Available: http://journal.umy.ac.id/index.php/jkm/article/view/212/174.
- [15] Y. Yuniati, "Pengaruh Berita di Surat Kabar terhadap Persepsi Mahasiswa tentang Politik," vol. 3, no. 1. pp. 1–17, 2002.
- [16] F. de Oliveira Capela and J. E. Ramirez-Marquez, "Detecting urban identity perception via newspaper topic modeling," *Cities*, vol. 93, no. May, pp. 72–83, 2019, doi: 10.1016/j.cities.2019.04.009.
- [17] X. Li, P. Wu, and W. Wang, "Incorporating stock prices and news sentiments for stock market prediction: A case of Hong Kong," *Inf. Process. Manag.*, vol. 57, no. 5, p. 102212, 2020, doi: 10.1016/j.ipm.2020.102212.
- [18] C. C. Lai, T. P. Shih, W. C. Ko, H. J. Tang, and P. R. Hsueh, "Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and coronavirus disease-2019 (COVID-19): The epidemic and the challenges," *Int. J. Antimicrob. Agents*, vol. 55, no. 3, p. 105924, 2020, doi: 10.1016/j.ijantimicag.2020.105924.
- [19] WHO, "Coronavirus disease 2019," World Heal. Organ., vol. 2019, no. March, p. 2633, 2020, doi: 10.1001/jama.2020.2633.
- [20] N. Reimers and I. Gurevych, "Optimal Hyperparameters for Deep LSTM-Networks for Sequence Labeling Tasks," 2017, [Online]. Available: http://arxiv.org/abs/1707.06799.
- [21] A. F. Hidayatullah, A. M. Hakim, S. Cahyaningtyas, and W. P. Aulia, "TOPIC MODELING DATA TWITTER TERHADAP CALON PRESIDEN REPUBLIK INDONESIA 2019 MENGGUNAKAN METODE LATENT DIRICHLET ALLOCATION (LDA)."
- [22] A. F. Hidayatullah, E. C. Pembrani, W. Kurniawan, G. Akbar, and R. Pranata, "Twitter Topic Modeling on Football News," 2018 3rd Int. Conf. Comput. Commun. Syst. ICCCS 2018, pp. 94–98, 2018, doi: 10.1109/CCOMS.2018.8463231.

Chairullah Naury, Copyright © 2021, MIB, Page 33 Submitted: **28/10/2020**; Accepted: **25/11/2020**; Published: **22/01/2021**