ANALISIS PERCEPATAN PEMULIHAN EKONOMI INDONESIA PASCA PANDEMI DENGAN BIG DATA DAN DEEP LEARNING

Rina [1], Taopik Hidayat [2], Daniati Uki Eka Saputri [3]
[1,3] Sistem Informasi, Universitas Nusa Mandiri
[2] Sains Data, Universitas Nusa Mandiri
Jl. Raya Jatiwaringin, Cipinang Melayu, Kec. Makasar, Jakarta Timur, DKI Jakarta 13620
taopik.toi@nusamandiri.ac.id

ABSTRAK

Penelitian ini membahas tentang percepatan pemulihan ekonomi Indonesia pasca pandemi COVID-19 melalui pendekatan analisis big data dengan penerapan teknik machine learning dan deep learning. Dengan munculnya pandemi pada akhir 2019, dampaknya menyebar ke seluruh dunia dan memicu serangkaian kebijakan, termasuk Pembatasan Sosial Berskala Besar (PSBB) di Indonesia. Kebijakan ini, sementara membantu menanggulangi penyebaran virus, namun secara signifikan menghentikan aktivitas ekonomi, terutama di sektor usaha kecil dan menengah (UMKM). Studi sebelumnya menyoroti dampak ekonomi global, kinerja keuangan di bursa efek Indonesia, dan bahkan implikasinya terhadap UMKM. Penelitian ini bertujuan mengkaji upaya percepatan pemulihan ekonomi di Indonesia pasca pandemi dengan memanfaatkan data teks dari berita dan ulasan pengguna hotel. Dengan menggunakan algoritma machine learning seperti Naive Bayes, Support Vector Machine, dan Random Forest, serta algoritma deep learning seperti Attention Mechanism dan Bidirectional LSTM, penelitian ini berusaha menghasilkan wawasan mendalam tentang tren dan pola perilaku masyarakat pasca pandemi. Data diperoleh dari sumber publik online dan diharapkan dapat memberikan kontribusi bagi pemangku kepentingan dalam merancang strategi pemulihan ekonomi yang efektif. Melalui analisis data teks yang komprehensif, penelitian ini diharapkan dapat memberikan pandangan tren dan pola perilaku dari data yang besar untuk memahami dinamika pemulihan ekonomi Indonesia dan memandu kebijakan yang lebih tepat sasaran. Hasil penelitian ini memperlihatkan tingkat akurasi masing-masing algoritma: SVM 88%, MultinomialNB 78%, Random Forest 84%, dan Bidirectional LSTM 92%. Analisis dilakukan pada data ulasan pengguna Traveloka, menunjukkan lonjakan signifikan dalam ulasan aplikasi. Hal ini mengindikasikan peningkatan penggunaan aplikasi, mencerminkan masyarakat yang mulai berpergian pasca pandemi, berpotensi mempengaruhi ekonomi Indonesia.

Kata kunci: *Pemulihan Ekonomi, Pemodelan Topik, Covid-19, Machine Learning, Deep Learning*

1. PENDAHULUAN

Pada akhir Desember 2019, ditemukan laporan sebanyak 27 kasus pneumonia atau infeksi paru-paru yang belum diketahui penyebabnya. Kasus ini pertama kali ditemukan di Kota Wuhan, Provinsi Hubei, China [1]. Beberapa gejala yang ditunjukkan oleh pasien yang mengalami infeksi tersebut di antaranya, batuk kering, demam, sesak nafas, dan lain sebagainya. Dengan gejala-gejala tersebut, penyakit ini menjadi wabah yang sangat cepat menular melalui rantai penularan antar manusia sehingga pada 30 Januari 2020, WHO mendeklarasikan wabah tersebut merupakan kedaruratan kesehatan masyarakat yang meresahkan dunia [2].

Pada tanggal 11 Februari 2020, secara resmi WHO menyebut penyakit ini merupakan penyakit yang dipicu oleh 2019-nCoV [3]. Per 31 Maret 2020, menurut data Worldmeters terkonfirmasi kasus positif sebanyak 801.117, pasien meninggal dunia sebanyak 38.771, dan 172.319 kasus dinyatakan sembuh pada 204 negara, salah satunya Indonesia. Pada 2 Maret 2020, pemerintah Indonesia untuk pertama kalinya mengumumkan pasien yang positif Covid-19. Pasca pengumuman tersebut, jumlah penyebaran kasus Covid-19 meningkat secara pesat setiap harinya [4].

Virus Covid-19 banyak menyita perhatian dunia karena telah memberikan dampak nagatif di berbagai bidang kehidupan, seperti pendidikan, kesehatan hingga perekonomian global [3]. Berbagai dampak pandemi terhadap perekonomian sudah dijelaskan oleh para peneliti sebelumnya. Pandemi Covid-19 telah berdampak pada perekonomian global [5], kinerja keuangan di Bursa Efek Indonesia [6], perbankan di Indonesia [1], bahkan berdampak pada usaha kecil dan menengah di Indonesia [3].

Pemerintah Indonesia menerapkan berbagai kebijakan untuk mencegah penyebarannya. Salah satu upaya yang diterapkan pemerintah Indonesia yaitu pembatasan sosial berskala besar (PSBB). Upaya ini meliputi pembatasan sekolah, tempat kerja, tempat peribadatan, transportasi dan tempat umum lainnya. Menurut [7], usaha kecil dan menengah (UMKM) sebagai salah satu sektor terdepan yang mengalami guncangan ekonomi karena pandemi Covid-19. Hal ini disebabkan adanya pembatasan sosial yang menghentikan aktivitas ekonomi secara tiba-tiba sehingga menghambat rantai pasokan di seluruh dunia.

Tujuan penelitian ini adalah untuk mengkaji upaya percepatan pemulihan ekonomi di Indonesia pasca pandemi melalui pendekatan analisis big data menggunakan teknik machine learning dan deep

learning. Dengan pandemi yang telah mengubah lanskap ekonomi, pemahaman mendalam mengenai cara terbaik untuk memulihkan ekonomi sangat penting.

Beberapa penelitian terdahulu membuktikan tingkat keberhasilan yang tinggi dalam analisis data teks, seperti prediksi kesehatan mental dengan menggunakan kombinasi word2vec dan BERT dengan Bi-LSTM mampu menganalisa data dari platform-media sosial populer Reddit dan Twitter, diperoleh hasil akurasi sebesar 98% dalam mendeteksi tanda-tanda depresi dan kecemasan dalam posting-media sosial. [8], penelitian lain yaitu model RNN dan LSTM diterapkan pada teks tweet terhadap enam layanan maskapai penerbangan Amerika. Hasil penelitian menunjukkan akurasi klasifikasi mencapai 80% [9], penelitian lain mengusulkan metode NLP untuk mengolah data dari sosial media untuk mendeteksi gangguan anxiety dan hasil mengindikasikan bahwa kesehatan mental di Indonesia yang dipengaruhi anxiety melesat jauh dari sebelumnya saat masapandemi belum berlangsung [10], dan penelitian lain mengusulkan metode MIDAS untuk membentuk model nowcasting pertumbuhan ekonomi Indonesia sebelum dan selama pandemi COVID-19, hasil diperoleh DataGoogle Trendsmampu memprediksipertumbuhanekonomi Indonesia sebelum pandemi lebih baik dari kelompok variabel lainnya. [11].

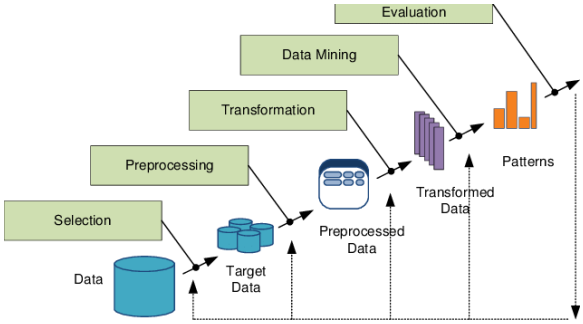
Penelitian lain memanfaatkan teknik analisa teks mining dilakukan oleh pamungkas & Iqbal tahun 2021, dengan menggunakan metode SVM, NaiveBayes, dan K-Nearest Neighbor untuk melihat hasil terbaik terhadap tanggapan masyarakat Indonesia terhadap pandemi Covid-19 pada media sosialTwitter. hasil menunjukkan model SVM paling baik [12].

Data untuk penelitian ini diperoleh dari sumber publik di internet, termasuk teks artikel berita berbahasa Indonesia dari situs-situs seperti kompas.com, cnbcindonesia.com, detik.com, viva.com, dan sindonews.com dengan kata kunci "covid-19". Selain itu, analisis juga akan dilakukan terhadap data ulasan pengguna hotel dari tahun 2020 hingga 2022 yang diperoleh dari aplikasi traveloka. Model analisis akan melibatkan algoritma machine learning seperti Naive Bayes, Support Vector Machine, dan Random Forest, serta algoritma deep learning seperti Attention Mechanism dan Bidirectional LSTM. Dengan menggabungkan analisis data teks dari berita dan ulasan pengguna, penelitian ini diharapkan dapat memberikan kontribusi signifikan bagi pemahaman pemulihan ekonomi di Indonesia pasca pandemi. Dengan memahami tren dan pola perilaku dari data yang besar, diharapkan strategi pemulihan ekonomi yang lebih efektif dapat dirancang

2. TINJAUAN PUSTAKA

2.1. Knowledge Discovery in Database (KDD)

Gambar 1 merupakan proses Knowledge Discovery in Database (KDD) yang dimulai dengan seleksi, di mana dataset relevan dipilih dari basis data yang besar untuk analisis tertentu, diikuti oleh pra-pemrosesan untuk meningkatkan kualitas data melalui penanganan nilai yang hilang, koreksi kesalahan, dan eliminasi duplikasi. Selanjutnya, transformasi dilakukan untuk mengubah data menjadi format yang lebih sesuai untuk mining, termasuk normalisasi dan pengurangan dimensi. Inti dari KDD adalah data mining, menggunakan teknik analitis untuk menemukan pola atau hubungan dalam data. Akhirnya, evaluasi dilakukan untuk menilai kegunaan dan validitas pengetahuan yang ditemukan, dengan memeriksa keakuratan, keandalan, dan relevansinya terhadap tujuan analisis awal, seringkali dengan feedback dari pengguna akhir atau ahli domain [13].



Gambar 1. Metode Knowledge Discovery in Database[13]

2.2. Case folding

Case folding merupakan konversi teks atau karakter ke bentuk yang tidak mempertimbangkan perbedaan huruf besar dan kecil. Ini melibatkan pengubahan seluruh huruf dalam suatu teks menjadi huruf kecil atau besar, memudahkan perbandingan dan pencarian tanpa memedulikan perbedaan ukuran huruf [14]. Proses ini umum digunakan dalam pengolahan teks, pencarian informasi, seperti basis data, mesin pencari, dan pemrosesan teks, untuk memastikan konsistensi dan keseragaman saat memanipulasi atau mencari data yang ditulis dengan huruf besar, huruf kecil, atau kombinasi keduanya.

2.3. Remove special character

Langkah untuk menghilangkan karakter khusus atau simbol dari suatu teks atau data. Karakter khusus ini mencakup simbol, tanda baca, atau karakter non-alfanumerik yang tidak termasuk huruf atau angka [15]. Penghapusan karakter khusus dilakukan untuk berbagai tujuan, seperti membersihkan teks sebelum menganalisis data, menjamin keseragaman format, atau menyiapkan data untuk pengolahan lebih lanjut. Dengan menghapus karakter khusus, teks menjadi lebih mudah diolah atau diinterpretasi dalam konteks tertentu.

2.4. Tokenizing

Proses membagi teks atau kalimat menjadi unit-unit lebih kecil yang disebut token. Token dapat berupa kata, frasa, simbol, atau unsur lainnya,

bergantung pada aturan tokenisasi yang digunakan. Tujuan dari tokenisasi adalah menyederhanakan teks agar dapat diolah atau dianalisis lebih efisien [16]. Dalam pemrosesan bahasa alami atau analisis teks, tokenisasi menjadi langkah kunci untuk memahami struktur dan makna teks. Melalui tokenisasi, komputer dapat dengan lebih baik memahami dan bekerja dengan teks dalam format yang lebih terstruktur.

2.5. Remove stopwords

Langkah untuk mengeliminasi kata-kata pengisi atau umum yang sering muncul dalam teks dan dianggap kurang berkontribusi pada informasi utama dalam konten tersebut [16]. Kata-kata ini, dikenal sebagai stopwords, biasanya tidak memberikan nilai tambah dalam analisis teks atau pemahaman kontennya. Contoh stopwords dalam bahasa Inggris termasuk "the", "and", "is", dan sejenisnya. Tindakan menghilangkan stopwords membantu menyederhanakan teks dan meningkatkan kualitas analisis dengan menitikberatkan pada kata-kata kunci atau makna yang lebih signifikan.

2.6. Stemming

Stemming merupakan langkah linguistik dalam pemrosesan bahasa alami yang bertujuan menghapus afiks (akhiran, awalan, atau imbuhan) dari kata-kata untuk menyisakan bentuk dasar atau akar kata. Tujuannya adalah menggabungkan kata-kata dengan akar kata yang sama, walaupun memiliki bentuk morfologis yang berbeda [17]. Sebagai contoh, kata-kata seperti "berlari", "berlari-lari", dan "berlarian" dapat disederhanakan menjadi bentuk dasar "lari" melalui proses stemming. Stemming membantu menyederhanakan analisis teks dan meningkatkan keseragaman dalam pemrosesan bahasa alami.

2.7. Term-Matrix (DTM)

DTM, singkatan dari Term-Matrix, merupakan representasi numerik dari kumpulan dokumen. Dalam DTM, setiap baris menggambarkan satu dokumen, dan setiap kolom mencerminkan kata-kata atau istilah unik dalam kumpulan tersebut. Nilai di setiap sel mencerminkan frekuensi atau bobot suatu kata dalam dokumen tertentu [18]. Penerapan DTM sangat umum dalam analisis teks dan pemrosesan bahasa alami, menyediakan landasan untuk menggunakan teknik seperti analisis sentimen, klasifikasi dokumen, dan pemodelan tema. DTM secara efektif mengubah teks menjadi representasi matematis, memungkinkan analisis lebih lanjut.

2.8. Pemodelan Topik (Latent Dirichlet Allocation)

Latent Dirichlet Allocation (LDA), merupakan cara untuk mengungkap topik atau tema dalam suatu koleksi dokumen. Dalam LDA, dokumen dianggap sebagai kombinasi dari beberapa topik, dan setiap kata dihasilkan oleh satu topik. Fokusnya adalah mengidentifikasi sebaran topik dalam koleksi dokumen dan distribusi kata dalam setiap topik [19]. LDA memungkinkan representasi dokumen sebagai campuran topik dengan bobot tertentu, bermanfaat dalam analisis teks, klasifikasi dokumen, dan pemahaman struktur tema dalam pemrosesan bahasa alami dan penambangan data.

2.9. Pelabelan Data (Lexicon Based)

Cara memberikan label atau mengklasifikasi data, terutama dalam analisis sentimen atau pemrosesan teks. Pendekatan ini memanfaatkan daftar kata kunci atau kamus (lexicon) yang telah terkategori, seperti positif atau negatif. Saat menganalisis teks, sistem mencocokkan kata-kata dalam teks dengan daftar kata kunci tersebut untuk menetapkan sentimen atau klasifikasi yang sesuai [20]. Pendekatan ini mengandalkan kamus kata kunci sebelumnya untuk memberi label pada data.

2.10. Machine learning

Machine learning merupakan subdisiplin kecerdasan buatan yang difokuskan pada pengembangan algoritma dan model komputer yang mampu belajar pola dari data dan mengambil keputusan atau tindakan tanpa perlu instruksi eksplisit [21]. Pendekatan ini memungkinkan mesin untuk meningkatkan kinerjanya seiring waktu dengan pengalaman dan pertambahan data. Aplikasi machine learning melibatkan berbagai bidang, termasuk pengenalan wajah, klasifikasi teks, prediksi, dan otomasi tugas-tugas kompleks.

2.11. Deep Learning

Deep learning adalah bagian dari machine learning yang fokus pada penerapan struktur jaringan saraf tiruan yang kompleks untuk memahami dan merepresentasikan data. Dalam deep learning, jaringan saraf terdiri dari banyak lapisan, memungkinkan model untuk secara otomatis mengekstrak fitur-fitur tingkat tinggi dari data tanpa perlu ekstraksi fitur manual [22]. Keberhasilan deep learning terutama terlihat dalam prestasi tinggi dalam tugas-tugas seperti pengenalan gambar, pemrosesan bahasa alami, dan tugas-tugas kompleks lainnya. Model deep learning dapat mengembangkan representasi hierarkis dari data, memungkinkannya menangani masalah-masalah yang sulit dan kompleks.

3. METODE PENELITIAN

Metode yang digunakan pada penelitian menggunakan model Knowledge Discovery in Database (KDD) yaitu proses mengekstraksi informasi baru dan pengetahuan dari database berukuran besar. Proses ekstraksi informasi dan pengetahuan diawali dengan pengumpulan data dari situs berita nasional yang akan digunakan untuk membangun pemodelan topik, model machine learning, dan deep learning. Selain itu, untuk menguatkan hasil penelitian dilakukan juga analisis terhadap data ulasan pengguna hotel mulai dari tahun 2020 hingga tahun 2022 yang

bersumber dari aplikasi traveloka. Analisis ini digunakan untuk mengulas pergerakan bisnis penginapan selama pandemi covid-19.

3.1. Pengumpulan Data

Penelitian ini menggunakan data teks hasil scraping dari 5 situs berita online nasional menggunakan keyword "covid-19" dan data ulasan pengguna pada aplikasi traveloka periode tahun 2020 hingga 2022. Di antara situs berita nasional yang dimaksud yaitu kompas.com, cnbcindonesia.com, detik.com, viva.com, dan sindonews.com. Salah satu alasan pemilihan 5 situs berita ini karena termasuk ke dalam daftar situs berita terpopuler [23]. Data yang terkumpul dari 5 situs tersebut sebanyak 20.003 artikel dan data ulasan pengguna dari aplikasi Traveloka sebanyak 104.141 ulasan. Penjelasan setiap atribut pada dataset yang digunakan, dapat dilihat pada tabel-tabel berikut:

Tabel 1. Atribut Dataset Berita

Fitur	Tipe Data	Deskripsi
Headline	Objek	Judul berita
Date	Objek	Tanggal publish berita
Url	Objek	Alamat situs berita
Content	Objek	Isi berita

Tabel 1 merupakan atribut yang dimiliki oleh dataset berita, terdiri dari atribut headline, date, url, dan content dengan tipe data dari keseluruhan atribut sama.

Tabel 2. Atribut Dataset Ulasan Aplikasi Traveloka

Fitur	Tipe Data	Deskripsi
Hotel *id*	*Float*	*Id* penginapan
Hotel *link*	Objek	Alamat situs penginapan
Nama hotel	Objek	Nama penginapan
Lokasi	Objek	Domisili penginapan
Provinsi	Objek	Provinsi tempat penginapan
Tipe	Objek	Jenis penginapan
Date	*Datetime*	Waktu pemberian *rating* dan ulasan
Nama *user*	Objek	Nama pengguna akun Traveloka
Rating	*Float*	*Rating* yang diberikan pengguna
Review	Objek	Ulasan yang diberikan pengguna

Tabel 2 merupakan atribut dataset ulasan aplikasi traveloka dengan jumlah 10 yang masing-masing penjelasan terlihat pada table di atas.

3.2. Pra-pemrosesan Data

Preprocessing data teks dapat dipahami sebagai proses pengubahan data tidak terstruktur menjadi data yang lebih terstruktur melalui beberapa tahapan meliputi case folding, removing special characters, tokenization, remove stopwords (filtering) dan stemming. Langkah ini dilakukan agar dapat menghasilkan data yang lebih baik sehingga dapat digunakan pada proses selanjutnya [4]. Berikut ini adalah langkah-langkah dari pra- pemrosesan data teks:

a. Case folding: adalah teknik yang digunakan untuk mengubah teks menjadi bentuk yang sama. Semua teks yang tersusun dari huruf kapital dan huruf kecil akan diubah menjadi huruf kecil (lower case).
b. Remove special character: digunakan untuk menghapus karakter spesial seperti tanda baca, angka, dan karakter kosong (spasi) yang tidak relevan.
c. Tokenizing: digunakan untuk memecah teks menjadi bagian-bagian yang lebih kecil (token).
d. Remove stopwords: adalah teknik yang digunakan untuk menghapus kata-kata yang muncul pada suatu bahasa tertentu dengan jumlah besar dan dianggap tidak memiliki makna yang terlalu penting. Beberapa contoh stopwords dalam bahasa Indonesia di antaranya "yang", "di", "ini", "dan", "dari", "tapi" dan lain sebagainya.
e. Stemming: adalah teknik mengubah kata berimbuhan menjadi kata dasarnya.

Tabel 3. Contoh Dataset Berita Sebelum dan Sesudah Pra-pemrosesan Data

Content	Content Clean
Jakarta, CNBC Indonesia - Para Menteri ….	jakarta cnbc indonesia para menteri….
Jakarta, CNBC Indonesia - Investasi di ….	jakarta cnbc indonesia investasi di ….
Jakarta, CNBC Indonesia - Peneliti di ….	jakarta cnbc indonesia peneliti di ….

Pada table 3 merupakan beberapa contoh conten berita yang belum dilakukan pre-processing dan sesudah dilakukan proses pre-processing.

3.3. Pembentukan Document Term Matrix

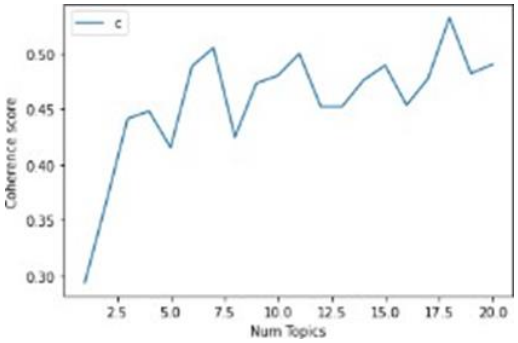
Setelah data berhasil dilakukan preprocessing, langkah selanjutnya yaitu menciptakan Document Term-Matrix (DTM). DTM adalah suatu matriks yang mencerminkan hubungan antara dokumen dengan kata atau istilah (term). Nilai DTM merupakan nilai yang dihasilkan dari Term Frequency-Inverse Document Frequency (TF-IDF). Term Frequency (TF) dimanfaatkan untuk mengetahui banyaknya (frekuensi) kata atau token yang terdapat dalam suatu dokumen. Sedangkan, Inverse Document Frequency (IDF) merupakan frekuensi kemunculan kata pada keseluruha dokumen.

3.4. Pemodelan Topik (Latent Dirichlet Allocation)

Pemodelan topik merupakan pemrosesan data berbasis statistik untuk menemukan sekumpulan topik dalam kumpulan dokumen teks tertentu. Alasan utama

menggunakan pemodelan topik adalah untuk mengungkapkan topik tersembunyi dalam kumpulan data teks yang besar. Terdapat beberapa teknik yang dapat digunakan untuk pemodelan topik, di antaranya LDA, LSA, PSA, NMF, dan lain sebagainya.

Pada penelitian ini, teknik yang digunakan adalah Latent Dirichlet Allocation (LDA). Teknik ini mengasumsikan bahwa sebuah dokumen dihasilkan berdasarkan sejumlah topik tertentu, dan setiap kata dalam dokumen dipilih secara acak dari kosakata topik yang sesuai [24]. Untuk menentukan banyaknya model topik yang dihasilkan dalam kumpulan dokumen teks dapat dilakukan dengan menghitung nilai coherence. Semakin tinggi nilai coherence yang dihasilkan, maka topik yang dihasilkan semakin baik. Berikut, grafik nilai coherence dataset berita yang digunakan pada penelitian ini:



Gambar 2. Nilai Coherence

Berdasarkan grafik pada gambar 2, terlihat nilai coherence tertinggi dihasilkan pada topik yang berjumlah 18 dengan nilai coherence sebanyak 0,53. Banyaknya topik yang dihasilkan pada tahap ini, akan digunakan sebagai acuan ketika membuat pemodelan topik pada tahap selanjutnya.

3.5. Pelabelan Data (Lexicon Based)

Sebelum melanjutkan pada tahap pembuatan pemodelan topik dan model klasifikasi menggunakan machine learning dan deep learning, terlebih dahulu dilakukan pelabelan data pada dataset berita yang digunakan. Pada penelitian ini, metode lexicon based digunakan untuk pelabelan data, Library Vader dipilih untuk mengelompokkan teks berdasarkan kamus lexicon sehingga menghasilkan class sentimen berupa negatif, netral, dan positif beserta tambahan compound score atau skor total. Nilai compound berfungsi sebagai satuan standar untuk mengklasifikasikan kalimat berdasarkan ketentuan class positif dengan compound $\geq 0{,}05$, negatif dengan compound $\leq -0{,}05$ dan netral dengan compound $> -0{,}05$ [12]. Di bawah ini merupakan hasil pelabelan yang dilakukan dengan menggunakan metode lexicon based:

Tabel 4. Hasil Pelabelan Data

Sentimen	Total
Positif	14.266
Netral	293
Negatif	5.444

Pada tabel 4 merupakan jenis sentiment yang dihasilkan terdiri dari sentiment positif yang berjumlah 14.266, sentiment netral berjumlah 293 dan sentiment negatif sejumlah 5.444.

3.6. Pemodelan Machine learning dan Deep Learning

Selain membuat pemodelan topik, penelitian ini juga menggunakan pendekatan machine learning. Terdapat beberapa tugas yang dapat dilakukan oleh machine learning yaitu klasifikasi, klasterisasi, prediksi, dan regresi. Namun, pada penelitian ini hanya akan digunakan satu tugas saja yaitu klasifikasi. Klasifikasi merupakan jenis machine learning yang bersifat supervised learning (pembelajaran terawasi) yang membutuhkan class. Class pada dataset berita dihasilkan melalui proses pelabelan data yang telah dilakukan sebelumnya. Pemilihan teknik klasifikasi pada penelitian ini bertujuan untuk mengklasifikasikan apakah sebuah berita termasuk ke dalam sentimen negatif atau positif berdasarkan data latih terhadap class yang sudah ditentukan.

Selain memanfaatkan pendekatan machine learning, penelitian ini juga megimplementasikan pendekatan Deep Learning. Algoritma Bidirectional Long ShortTerm Memori (LSTM) dipilih untuk melakukan klasifikasi pada dataset berita pandemi covid-19.

4. HASIL DAN PEMBAHASAN

4.1. Pemodelan Topik

Untuk membangun pemodelan topik menggunakan metode LDA, penelitian ini menggunakan package Gensim dengan melakukan pemanggilan modul genism.model. Banyaknya topik yang akan dibuat didasarkan pada hasil perhitungan nilai coherence yang telah dilakukan pada proses sebelumnya yaitu sebanyak 18 topik. Berikut ini 5 topik teratas yang dihasilkan dari dataset berita pandemi covid-19:

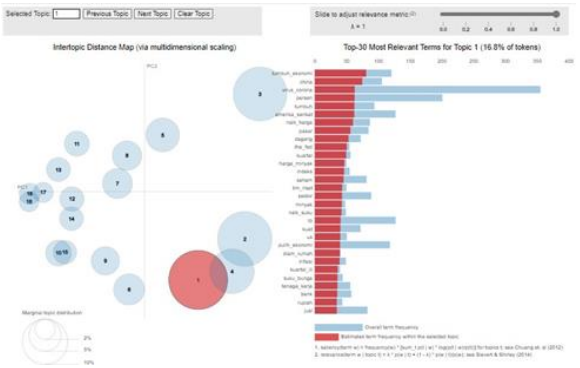
Tabel 5. Hasil Topik Modeling

Topik	Kata
1	tumbuh_ekonomi,china,virus_corona, persen,tumbuh, amerika_serikat, naik_harga, pasar, dagang, the fed
2	ma_ruf, virus_corona, Jokowi, dicky, arab_saudi, pulih_ekonomi, ma, tanam, ruf, persen
3	dki_jakarta, virus_corona, rumah_sakit, pasien, tinggal_dunia, dki, protokol_sehat, pasien_sembuh, menteri_sehat, jawa
4	virus_corona, kota_bekas, anak, laksana_pilkada, olahraga, social_distancing, persen, korea_selatan, minus_persen, sepak_bola
5	virus_corona, anak_anak, varian, omicron, varian_omicron, anak, vaksin, infeksi, sars_cov, who

Dari hasil pemodelan LDA pada table 5 di atas diperoleh informasi tema pemberitaan selama pandemi covid-19 pada 5 situs berita online yang digunakan pada penelitian ini adalah:

a. Topik 1: Seputar pertumbuhan ekonomi Indonesia selama pandemi dan kenaikan harga barang.
b. Topik 2: Seputar pemerintah dan pemulihan ekonomi saat pandemi.
c. Topik 3: Seputar informasi warga yang sembuh dan meninggal karena virus corona pada provinsi DKI Jakarta dan Jawa.
d. Topik 4: Seputar kebijakan social distancing selama pandemi dan pelaksanaan pilkada serta olah raga sepak bola.
e. Topik 5: Seputar varian virus corona dan vaksin

Untuk mempermudah dalam memahami struktur kata yang menyusun setiap topik, model topik yang dihasilkan kemudian divisualisasikan menggunakan library PyLDAvis. Hasil visualisasi ini menyediakan dua sisi panel (gambar 3). Sisi kiri memperlihatkan jumlah topik secara keseluruhan yaitu 20 topik dan sisi kanan menunjukan distribusi frekuensi token atau kata yang muncul di setiap topik yang dihasilkan. Pada panel sebelah kana, bar chart berwarna biru menunjukan term frequency secara keseluruhan dalam corpus. Sedangkan warna merah mengindikasikan estimasi term frequency pada topik yang dimaksud. Selain itu, pada sisi kiri terlihat beberapa topik memiliki jarak yang sangat dekat hingga tumpang tindih, hal ini menunjukkan bahwa terdapat kata-kata serupa yang membangun topik yang berbeda.



Gambar 3. Visualisasi Pemodelan Topik

4.2. Pemodelan machine learning dan Deep Learning

Pada tahap pembuatan model machine learning, beberapa algoritma seperti MultinomialNB, Support Vector Machine, dan Random Forest dipilih untuk melakukan klasifikasi pada dataset berita yang telah dilakukan pelabelan pada tahap sebelumnya. Sedangkan pembuatan model deep learning menggunakan Attention Mechanism with Bidirectional LSTM. Data yang memiliki label atau class netral, tidak akan digunakan dalam proses klasifikasi karena jumlah datanya yang sedikit, sehingga proses klasifikasi hanya menggunakan data yang memiliki label atau class positif dan negatif saja.

Setelah itu, hasil pengolahan data tersebut akan menghasilkan nilai akurasi. Kemudian, hasil evaluasi tersebut akan dikomparasi agar diketahui algoritma mana yang menghasilkan tingkat akurasi paling tinggi di antara ketiganya. Langkah berikutnya, dilakukan pembuatan model klasifikasi dengan menggunakan dataset yang telah dilakukan penyeimbangan menggunakan metode SMOTE. Sedangkan pada deep learning, penyeimbangan data menggunakan teknik loss function. Hal ini dilakukan karena dataset yang digunakan tidak seimbang (imbalance) dan algoritma SVM rentan terhadap dataset yang imbalance. Penyeimbangan data penting dilakukan jika class pada dataset yang digunakan imbalance. Hal ini disebabkan karena akurasi tersebut bisa saja hanya bernilai tinggi pada class yang bernilai dominan namun tidak pada data yang memiliki class yang sedikit.

Tabel 6. Dataset sebelum dan sesudah Over Sampling

Sentiment	Before Over Sampling	After Over Sampling
Positif	11403	11403
Negatif	4365	11403

Berdasarkan model machine learning dan deep learning yang telah dibangun pada tahap sebelumnya, hasil sebelum dan sesudah over sampling ditunjukkan pada table 6 di atas.

Tabel 7. Perbandingan Kinerja Algoritma Sebelum Penyeimbangan Data

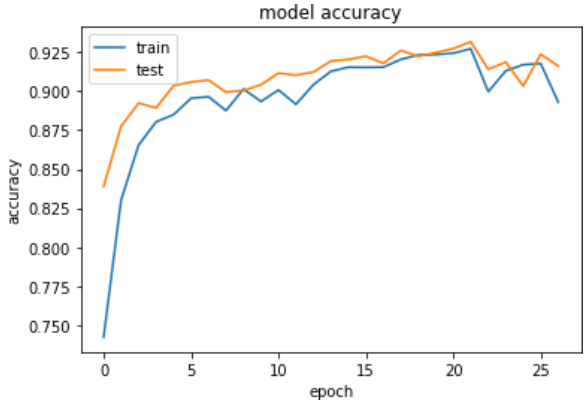
Algoritma	Akurasi
MultinomialNB	76%
Support Vector Machine	88%
Random Forest	**83%**

Table 7 merupakan hasil akurasi dari masing-masing algoritma, terlihat bahwa akurasi tertinggi diperoleh dari random forest sebesar 83% kemudian diikuti akurasi terbaik kedua yaitu support vector machine sebesar 88%, dan terakhir algoritma MultinomialNB sebesar 76%.

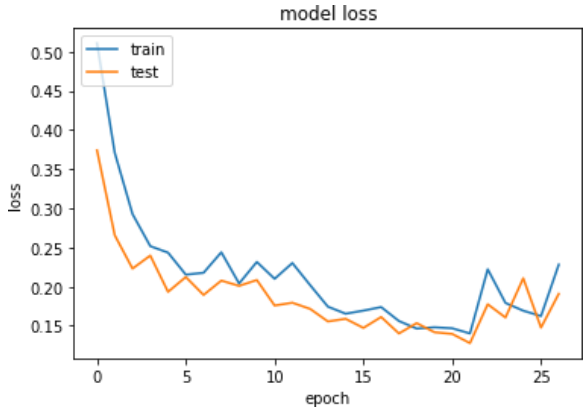
Tabel 8. Perbandingan Kinerja Algoritma Setelah Penyeimbangan Data

Algoritma	Accuracy
MultinomialNB + SMOTE	78%
Support Vector Machine + SMOTE	88%
Random Forest + SMOTE	84%
Bidirectional LSTM	92%

Pada tabel 8 di atas merupakan hasil perbandingan kinerja algoritma setelah penyeimbangan data. Dari keseluruhan hasil evaluasi di atas, terlihat akurasi pada algoritma MultinomialNB dan Random Forest mengalami kenaikan setelah dilakukan penyeimbangan data menggunakan SMOTE. Namun, akurasi pada algoritma SVM terlihat tidak mengalami perubahan. Sedangkan pada deep learning, setelah digunakan teknik penyeimbangan data menggunakan lost function, menjadikannya algoritma yang memiliki akurasi tertinggi dibanding algoritma lainnya.

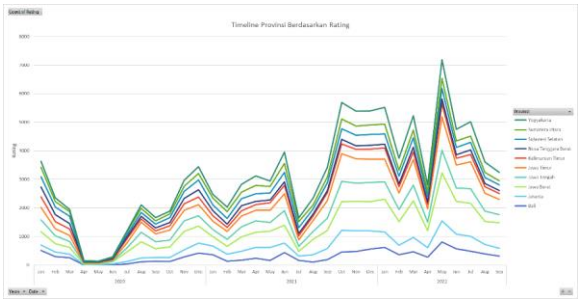


Gambar 4. Performa Model Accuracy



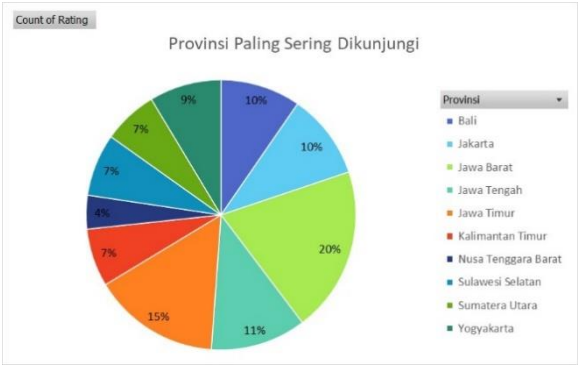
Gambar 5. Performa Model Loss

Pada gambar 4 dan gambar 5 merupakan hasil performa model deep learning yang ditampilkan dalam bentuk grafik. Terlihat bahwa ketika model diberikan parameter epoch sebesar 25 menunjukkan hasil akurasi yang semakin membaik, dan sebaliknya nilai kesalahan semakin mengecil.



Gambar 6. Trend Pemberian Rating pada Aplikasi Traveloka

Terlihat pada gambar 6 merupakan timeline pemberian rating di atas menunjukan pada bulan April dan Mei 2020 cenderung tidak ada pemberian rating yang dilakukan oleh pengguna. Hal ini disebabkan karena adanya kebijakan PSBB yang dilakukan oleh pemerintah Indonesia. Namun, pada bulan Mei 2022 pemberian rating oleh user pada seluruh jenis penginapan di 10 Provinsi mengalami kenaikan yang signifikan. Hal ini bertepatan dengan adanya libur Idul Fitri dan kebijakan pemerintah yang memperbolehkan masyarakat untuk mudik. Informasi lain yang diperoleh dari analisis ulasan pengguna aplikasi Traveloka dapat diketahui beberapa provinsi yang paling banyak dikunjungi selama 3250andemic. Grafik pie chart pada gambar 7 menunjukan provinsi Jawa Barat adalah provinsi yang paling banyak dikunjungi dengan presentase sebanyak 20%, kemudian diikuti provinsi Jawa Timur sebanyak 15%, dan seterusnya. Deskripsi lengkap dapat dilihat pada gambar berikut:



Gambar 7. Provinsi Paling Sering Dikunjungi

Pada gambar 7 merupakan persentase provinsi paling sering dikunjungi, terlihat bahwa 3 provinsi yang sering dikunjungi adalah Jawa Barat, Jawa Timur, dan Jakarta, sedangkan provinsi yang jarang dikunjungi adalah Nusa Tenggara Barat.

5. KESIMPULAN DAN SARAN

Penelitian menggunakan pendekatan analisis big data dan model machine learning serta deep learning terhadap data berita pandemi di Indonesia yang diperoleh melalui 5 situs berita online sebanyak 20.003 artikel. Setelah dilakukan pemodelan, didapatkan hasil akurasi dari masing- masing algoritma yaitu SVM 88%, MultinomialNB 78%, Random Forest 84%, dan Bidirectional LSTM 92%. Selain itu, analisis dilakukan terhadap data ulasan pengguna di aplikasi Traveloka. Hasil menunjukan bahwa terjadi kenaikan signifikan terhadap pemberian ulasan di aplikasi tersebut. Dengan kata lain, terjadinya lonjakan kenaikan pengguna pada aplikasi tersebut mengindikasikan bahwa masyarakat mulai bepergian ke lokasi lain dan secara tidak langsung berdampak pada perekonomian Indonesia pasca pandemi.

DAFTAR PUSTAKA

[1] Cakranegara, P. A. Effects of Pandemic Covid 19 on Indonesia Banking : *Ilomata International Journal of Management*, *1*(4), Article 4. https://doi.org/10.52728/ijjm.v1i4.161. 2020.

[2] Lu, Y., Zhao, J., Wu, X., & Lo, S. M. Escaping to nature during a pandemic: A natural experiment in Asian cities during the COVID-19 pandemic with big social media data. *Science of The Total Environment*, *777*, 146092. https://doi.org/10.1016/j.scitotenv.2021.146092. 2021.

[3] Mishra, N. P., Das, S. S., Yadav, S., Khan, W., Afzal, M., Alarifi, A., kenawy, E.-R., Ansari, M. T., Hasnain, M. S., & Nayak, A. K. Global impacts of pre- and post-COVID-19 pandemic: Focus on socio-economic consequences. *Sensors International*, *1*, 100042. https://doi.org/10.1016/j.sintl.2020.100042. 2020.

[4] Nurhadi, A., & Indrayuni, E. Sistem Informasi Pendaftaran Vaksinasi Covid-19. *JISICOM (Journal of Information System, Informatics and Computing)*, *5*(2), Article 2. https://doi.org/10.52362/jisicom.v5i2.491. 2021.

[5] Kohlscheen, E., Mojon, B., & Rees, D. . The Macroeconomic Spillover Effects of the Pandemic on the Global Economy. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3569554. 2020.

[6] Devi, S., Warasniasih, N. M. S., Masdiantini, P. R., & Musmini, L. S. The Impact of COVID-19 Pandemic on the Financial Performance of Firms on the Indonesia Stock Exchange. *Journal of Economics, Business, & Accountancy Ventura*, *23*(2), Article 2. https://doi.org/10.14414/jebav.v23i2.2313. 2020.

[7] Faizi, F., Wulandana, N. P., Alya, A., & Lombu, A. A. DAMPAK PANDEMI COVID-19 TERHADAP UMKM DI INDONESIA. *JURNAL LENTERA BISNIS*, *11*(2), Article 2. https://doi.org/10.34127/jrlab.v11i2.510. 2022.

[8] Zeberga, K., Attique, M., Shah, B., Ali, F., Jembre, Y. Z., & Chung, T.-S. A Novel. Text Mining Approach for Mental Health Prediction Using Bi-LSTM and BERT Model. *Computational Intelligence and Neuroscience*, *2022*, e7893775. https://doi.org/10.1155/2022/7893775. 2022.

[9] Monika, R., Deivalakshmi, S., & Janet, B. Sentiment Analysis of US Airlines Tweets Using LSTM/RNN. *2019 IEEE 9th International Conference on Advanced Computing (IACC)*, 92–95. https://doi.org/10.1109/IACC48062.2019.8971592. 2019.

[10] Putra V.P, Achmad Ryvaldy & Ahmad Rafi Syaifudin. Deteksi Anxiety Disorder Pengguna Sosial Media Menggunakan Deep Transfer Learning untuk Percepatan Pemulihan Pasca Pandemi. The Journal on Machine Learning and Computational Intelligence (JMLCI). e-ISSN: 2808-974X. 2023..

[11] Kurniawan, M., & Falentina, A. Analisis Big Data dan Official Statistics dalam Melakukan Nowcasting Pertumbuhan Ekonomi Indonesia Sebelum dan Selama Pandemi COVID-19. Seminar Nasional Official Statistics, 2022(1), 521-532. https://doi.org/10.34123/semnasoffstat.v2022i1.1146. 2022.

[12] Pamungkas, F. S., & Kharisudin, I. (2021, February). Analisis Sentimen dengan SVM, NAIVE BAYES dan KNN untuk Studi Tanggapan Masyarakat Indonesia Terhadap Pandemi Covid-19 pada Media Sosial Twitter. In PRISMA, Prosiding Seminar Nasional Matematika (Vol. 4, pp. 628-634).

[13] Chumbar Shawn. Knowledge Discovery in Databases (KDD): A Practical Approach. medium.com/@shawn.chumbar/knowledge-discovery-in-databases-kdd-a-practical-approach-f28247493be4. 24 September 2023.

[14] Rizal M, Martanto M, Hayati U. Analisis Sentimen Pengguna Twitter Terkait Film One Piece Menggunakan Metode Naive Bayes. Jurnal Sistem Informasi Kaputama (JSIK). 2024 Jan 1;8(1):38-47.

[15] Syafira, Firya. Analisis Sentimen Dampak Perkembangan Artificial Intelligence (AI) Pada Media Sosial Twitter Menggunakan Metode Support Vector Machine dan Lexicon Based. BS thesis. Fakultas Sains dan Teknologi UIN Syarif Hidayatullah Jakarta, 2023.

[16] Hidayat T, Pebrianto R, Pratiwi RL, Saputri DU. Implementasi Algoritma Klasifikasi Terhadap Tweet Pornografi Kaum Homoseksual Pada Twitter. Indonesian Journal on Software Engineering (IJSE). 2020 Dec 2;6(2):204-12.

[17] Jabbar A, Iqbal S, Tamimy MI, Rehman A, Bahaj SA, Saba T. An Analytical Analysis of Text Stemming Methodologies in Information Retrieval and Natural Language Processing Systems. IEEE Access. 2023 Nov 14;11:133681-702.

[18] Sapanji RV, Hamdani D, Harahap P. Sentiment Analysis of the Top 5 E-commerce Platforms in Indonesia using Text Mining and Natural Language Processing (NLP). Journal of Applied Informatics and Computing. 2023 Nov 30;7(2):202-11.

[19] Maulidiya, D., Topic Modelling using Latent Dirichlet Allocation (LDA) to Investigate the Latent Topics of Mathematical Creative Thinking Research in Indonesia. Journal of Intelligent Computing and Health Informatics (JICHI), 3(2), pp.35-46. 2023.

[20] Setiawan, H. and Zufria, I., Analisis Sentimen Pembatalan Indonesia Sebagai Tuan Rumah Piala Dunia FIFA U-20 Menggunakan Naïve Bayes. JURNAL MEDIA INFORMATIKA BUDIDARMA, 7(3), pp.1003-1012. 2023.

[21] Kühl, N., Schemmer, M., Goutier, M. and Satzger, G., Artificial intelligence and machine learning. Electronic Markets, 32(4), pp.2235-2244. 2022.

[22] Yessy Asri, S.T., Kuswardani, D. and Kom, M., MACHINE LEARNING & DEEP LEARNING: Analisis Sentimen Menggunakan Ulasan Pengguna Aplikasi. Uwais Inspirasi indonesia. 2024.

[23] Yana, M. D., & Setiawan, H. Analisis Framing Berita Perempuan Bakar Diri Dalam Media Online Detik.Com Dan Kompas.Com. *Jurnal Ilmiah Wahana Pendidikan*, *9*(25), Article 25. https://doi.org/10.5281/zenodo.10416729. 2023.

[24] Xie, R., Chu, S. K. W., Chiu, D. K. W., & Wang, Y. Exploring Public Response to COVID-19 on Weibo with LDA Topic Modeling and Sentiment Analysis. *Data and Information Management*, *5*(1), 86–99. https://doi.org/10.2478/dim-2020-0023. 2021.