

BAB II

TINJAUAN PUSTAKA

2.1 Landasan Teori

Pada sub-bab ini akan menjelaskan definisi dan karakteristik fundamental yang menjadi dasar pemahaman terkait penelitian yang dilakukan.

2.1.1. Berita Hoaks

Berita hoaks didefinisikan sebagai informasi palsu atau menyesatkan yang sengaja direkayasa untuk menipu atau memanipulasi pembaca [20]. Fenomena ini seringkali dicirikan oleh kecepatan penyebaran yang tinggi di platform digital, terutama media sosial, serta kemampuannya dalam memicu reaksi emosional yang kuat pada audiens [21]. Dampak berita hoaks sangat luas, mencakup disrupsi sosial dan politik, kerugian ekonomi, serta ancaman serius terhadap keamanan siber dan integritas informasi digital dalam ekosistem daring [22].

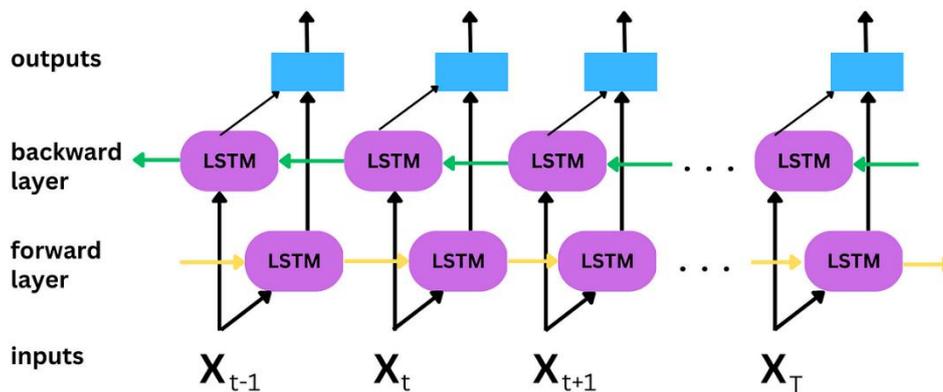
2.1.2. Pemrosesan Bahasa Alami (NLP)

NLP merupakan cabang dari *Artificial Intelligence* (AI) yang berfokus pada interaksi antara komputer dan bahasa manusia. Tujuan utama NLP adalah memungkinkan komputer untuk memahami, menafsirkan, dan menghasilkan bahasa manusia secara bermakna [23]. Dalam konteks analisis teks, NLP melibatkan berbagai tahapan seperti pembersihan teks dan ekstraksi fitur, [24] yang krusial untuk mengubah data tekstual menjadi format yang dapat diproses oleh model komputasi.

2.1.3. Pembelajaran Mendalam (Deep Learning)

DL merupakan evolusi dari *Machine Learning* (ML) yang menggunakan arsitektur jaringan saraf tiruan (*Neural Network*) dengan banyak lapisan (*deep layers*) untuk mempelajari representasi data secara

hierarkis [25]. Berbeda dengan ML tradisional, DL mampu mengekstraksi fitur kompleks secara otomatis dari data mentah, menjadikannya sangat efektif untuk tugas-tugaa seperti pengenalan gambar, suara, dan pemrosesan bahasa alami [26]. Jaringan saraf tiruan umumnya terdiri dari lapisan masukan (*input layer*), beberapa lapisan tersembunyi (*hidden layers*), dan lapisan keluaran (*output layer*) [25].



Gambar 2. 1. Ilustrasi BiLSTM

Sistematika Bi-LSTM pada dasarnya melibatkan pemrosesan sebuah sekuens input oleh dua lapisan LSTM seperti pada gambar 2.1 yang bekerja secara simultan dalam arah yang berlawanan. Lapisan pertama, yaitu *forward layer*, memproses data secara kronologis dari awal hingga akhir untuk menangkap konteks masa lalu dari setiap elemen dalam sekuens. Secara bersamaan, *backward layer* memproses data yang sama dari akhir ke awal untuk menangkap konteks masa depan. Pada setiap langkah waktu, hasil dari kedua lapisan ini—yaitu vektor tersembunyi dari arah maju dan vektor tersembunyi dari arah mundur—kemudian digabungkan untuk membentuk satu representasi vektor yang utuh. Melalui mekanisme ini, Bi-LSTM mampu menghasilkan pemahaman kontekstual yang jauh lebih kaya

untuk setiap elemen, karena memperhitungkan informasi dari kedua sisi (sebelum dan sesudah) secara bersamaan.

2.2 Pra-pemrosesan Teks

Pra-pemrosesan teks adalah tahapan penting dalam NLP yang bertujuan untuk membersihkan dan mengubah data teks mentah menjadi format yang sesuai untuk analisis dan pelatihan model.

2.2.1. Pembersihan Teks (Text Cleaning)

Pembersihan teks adalah proses untuk menghilangkan elemen-elemen yang tidak relevan atau mengganggu dari data teks agar model dapat fokus pada informasi yang bermakna [10]. Langkah-langkah umum dalam pembersihan teks meliputi konversi seluruh teks menjadi huruf kecil (*lowercase*), penghapusan Uniform Resource Locators (URL), tag HTML, tanda baca, angka, dan karakter khusus[27]. Normalisasi spasi berlebih juga dilakukan untuk memastikan konsistensi format teks[28].

2.2.2. Tokenisasi

Setelah pembersihan teks, tahap selanjutnya adalah tokenisasi. Tokenisasi adalah proses memecah urutan karakter teks menjadi unit-unit yang lebih kecil dan bermakna yang disebut "token" [29]. Dalam konteks pemrosesan bahasa, token umumnya berupa kata atau frasa pendek. Proses ini memungkinkan model untuk memproses teks sebagai daftar item diskrit, yang merupakan langkah awal penting sebelum analisis linguistik lebih lanjut seperti penghapusan stopwords atau stemming. Teknik tokenisasi dapat bervariasi, mulai dari pemisahan berdasarkan spasi hingga metode yang lebih kompleks yang mempertimbangkan aturan tata bahasa dan pengecualian.

2.2.3. Penghapusan Stopword

Penghapusan stopword adalah proses eliminasi kata-kata yang umum muncul dalam suatu bahasa namun memiliki nilai informatif yang rendah dalam analisis teks. Contoh stopword dalam bahasa Indonesia meliputi "yang", "dan", "di", "ke", "dari", dan sebagainya. Tujuan dari tahap ini adalah untuk mengurangi dimensi data dan meningkatkan efisiensi komputasi, serta membantu model untuk fokus pada kata-kata kunci yang lebih relevan dan memiliki bobot semantik yang lebih tinggi untuk tugas tertentu (misalnya, klasifikasi atau analisis sentimen)[30]. Daftar stopword biasanya berasal dari korpus atau kamus stopword yang sudah ditetapkan untuk bahasa tertentu.

2.2.4. Stemming

Stemming adalah proses morfologis yang bertujuan untuk mengurangi kata-kata berinfleksi atau turunan (misalnya, "bermain", "pemain", "dimainkan") menjadi bentuk dasar atau akar kata (root word) tanpa jaminan bahwa kata dasar tersebut adalah kata yang valid secara morfologis [31]. Meskipun tidak selalu menghasilkan kata dasar yang benar secara linguistik, stemming efektif dalam mengelompokkan berbagai bentuk kata yang memiliki makna dasar yang sama, sehingga mengurangi jumlah fitur unik dalam data dan membantu generalisasi model. Dalam konteks penelitian ini, stemming diterapkan khususnya pada data berbahasa Melayu menggunakan algoritma yang sesuai untuk bahasa tersebut.

2.3 Algoritma dan Model Pembelajaran Mendalam

Sub-bab ini menjelaskan secara rinci arsitektur *deep learning* yang menjadi inti dari model penelitian ini.

2.3.1 Long Short-Term Memory (LSTM)

Long Short-Term Memory (LSTM) adalah varian khusus dari Recurrent Neural Network (RNN) yang dirancang untuk mengatasi masalah *vanishing* atau *exploding gradient* yang sering terjadi pada RNN tradisional saat memproses dependensi jangka panjang dalam data sekuensial. LSTM diperkenalkan oleh Hochreiter dan Schmidhuber pada tahun 1997 dan dikenal memiliki "sel memori" yang memungkinkannya menyimpan informasi untuk periode waktu yang lama[32]. Struktur LSTM terdiri dari tiga gerbang utama: *forget gate* yang mengontrol informasi mana yang harus dibuang dari sel memori, *input gate* yang mengatur informasi baru mana yang akan ditambahkan ke sel memori, dan *output gate* yang menentukan bagian mana dari sel memori yang akan menjadi *output* [25].

2.3.2 Bidirectional Long Short-Term Memory (BiLSTM)

Bidirectional Long Short-Term Memory (BiLSTM) adalah pengembangan dari LSTM yang memproses informasi dalam dua arah: maju (dari awal ke akhir *sequence*) dan mundur (dari akhir ke awal *sequence*) [33]. Kemampuan ini memungkinkan BiLSTM untuk menangkap konteks yang lebih kaya dan utuh dari teks, karena dapat mempertimbangkan informasi dari masa lalu (sebelumnya) dan masa depan (setelahnya) dari suatu kata dalam *sequence* [16]. Keunggulan ini membuat BiLSTM sangat efektif untuk tugas-tugas klasifikasi teks seperti deteksi hoaks, di mana pemahaman konteks kalimat secara menyeluruh sangat krusial .

2.4 Optimisasi dan Evaluasi Model

Sub-bab ini mencakup aspek-aspek penting dalam pelatihan dan penilaian kinerja model.

2.4.1 Hyperparameter Tuning

Hyperparameter adalah parameter eksternal yang nilainya diatur sebelum proses pelatihan model dimulai dan tidak dipelajari dari data [34]. Contoh *hyperparameter* meliputi *learning rate*, ukuran *batch*, dan jumlah *epoch*. *Hyperparameter tuning* adalah proses mengoptimalkan nilai-nilai ini untuk menemukan kombinasi terbaik yang menghasilkan performa model yang optimal [35].

2.4.2 **Optimizer (Adam)**

Optimizer adalah algoritma atau metode yang digunakan untuk menyesuaikan atribut jaringan saraf, seperti bobot dan *learning rate*, dengan tujuan meminimalkan fungsi kerugian (*loss function*) selama proses pelatihan [34]. Penelitian ini menggunakan *Adaptive Moment Estimation* (Adam) sebagai *optimizer* [27]. Adam adalah algoritma optimisasi adaptif yang populer, dikenal karena efisiensinya dalam komputasi dan kemampuannya untuk mengoptimalkan *learning rate* untuk setiap parameter berdasarkan estimasi momen pertama dan kedua dari gradien [27].

2.4.3 **Fungsi Kerugian (Loss Function)**

Fungsi kerugian (*loss function*) mengukur seberapa baik model memprediksi *output* yang benar dengan menghitung selisih antara nilai prediksi model dan nilai aktual. Untuk tugas klasifikasi biner seperti deteksi hoaks, *Binary Cross-Entropy* (BCE) adalah fungsi kerugian yang umum digunakan [28]. BCE mengukur kinerja model klasifikasi yang *output*-nya adalah nilai probabilitas antara 0 dan 1 [36].

2.4.4 **Metrik Evaluasi**

Metrik evaluasi digunakan untuk mengukur kinerja model klasifikasi dalam memprediksi kelas target [29]. Untuk masalah

deteksi hoaks yang mungkin memiliki distribusi kelas tidak seimbang, beberapa metrik relevan meliputi Akurasi (*Accuracy*), Presisi (*Precision*), Rekal (*Recall*), F1-Score, Matriks Kebingungan (*Confusion Matrix*)[35].

2.4.4.1. Akurasi

Accuracy adalah metrik yang paling umum digunakan dan mengukur proporsi total prediksi yang benar (baik positif maupun negatif) dari total jumlah sampel. Akurasi di definisikan dengan rumus berikut:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (2.1)$$

Rumus 2. 1. Rumus Akurasi

*Penjelasan pada 2.4.4.5

2.4.4.2. Presisi

Precision (juga dikenal sebagai *Positive Predictive Value*) mengukur proporsi instans yang benar-benar positif dari semua instans yang diprediksi sebagai positif. *Precision* penting ketika biaya *false positive* (mengklasifikasikan berita non-hoax sebagai hoax) sangat tinggi. Presisi di rumuskan sebagai berikut:

$$Precision = \frac{TP}{TP+FP} \quad (2.2)$$

Rumus 2. 2. Rumus Presisi

*Penjelasan pada 2.4.4.5

2.4.4.3. Rekal

mengukur proporsi instans positif yang benar-benar terdeteksi oleh model dari semua instans positif yang sebenarnya ada. *Recall* penting ketika biaya *false negative* (mengklasifikasikan berita hoax sebagai non-hoax) sangat tinggi. Rekal di rumuskan sebagai berikut:

$$Recall = \frac{TP}{TP + FN} \quad (2.3)$$

Rumus 2. 3. Rumus Rekal

*Penjelasan pada 2.4.4.5

2.4.4.4. F1-score

F1-Score adalah rata-rata harmonik dari *precision* dan *recall* [16]. Metrik ini sangat berguna pada dataset yang tidak seimbang karena memberikan bobot yang seimbang pada *precision* dan *recall*, memberikan gambaran tunggal tentang kinerja model yang menggabungkan kedua aspek tersebut. F1-score di rumuskan sebagai berikut:

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (2.4)$$

Rumus 2. 4. Rumus f1-score

*Penjelasan pada 2.4.4.5

2.4.4.5. Matrik Kebingungan

Confusion Matrix adalah tabel ringkasan yang digunakan untuk memvisualisasikan kinerja algoritma klasifikasi pada set data uji, menunjukkan jumlah True

Positives (TP), True Negatives (TN), False Positives (FP), dan False Negatives (FN) [16].

- True Positives (TP): Jumlah prediksi positif yang benar (misalnya, berita hoax yang benar-benar hoax).
- True Negatives (TN): Jumlah prediksi negatif yang benar (misalnya, berita non-hoax yang benar-benar non-hoax).
- False Positives (FP): Jumlah prediksi positif yang salah (misalnya, berita non-hoax yang salah diklasifikasikan sebagai hoax, juga dikenal sebagai Type I error).
- False Negatives (FN): Jumlah prediksi negatif yang salah (misalnya, berita hoax yang salah diklasifikasikan sebagai non-hoax, juga dikenal sebagai Type II error).

Confusion matrix adalah dasar dari semua metrik evaluasi lainnya dan memberikan wawasan mendalam tentang jenis kesalahan yang dibuat model.

2.3 Penelitian Terdahulu

Tabel berikut menyajikan perbandingan hasil dari penelitian-penelitian sebelumnya yang menggunakan metode Bi-LSTM.

Tabel 2.1 Penelitian Terdahulu

No	Nama Jurnal, Vol, Nomor, Tahun	Judul Artikel, Penulis	Hasil
1	IEEE Access, 9, 2021.	Deep Neural Networks for Predicting Solar Radiation at Hail Region, Saudi Arabia, [37] Sahbi Boubaker, Mohamed Benganem, Adel Mellit, Ayoub	mengembangkan dan mengevaluasi berbagai model <i>deep neural network</i> (DNN) untuk memprediksi iradiasi matahari harian (GHI) satu hari ke depan di Hail, Arab Saudi. Model yang diuji meliputi LSTM, BiLSTM, GRU, Bi-GRU, CNN, serta model hibrida seperti CNN-

		Lefza, Omar Kahouli, dan Liqia Kolsi.	LSTM dan CNN-BiLSTM, dengan hanya menggunakan data historis GHI. Di antara semua model yang diuji, model Bi-LSTM menunjukkan <i>Mean Absolute Percentage Error</i> (MAPE) terendah yaitu sebesar 5,24% dengan akurasi 96%.
2	Procedia Computer Science (dari 2022 IEEE International Conference on Distributed Computing and Electrical Circuits and Electronics), 2022.	A Comparative Analysis of ARIMA, GRU, LSTM and BiLSTM on Financial Time Series Forecasting[38], Muskaan Pirani, Mohammed Husain Bohara, Paurav Thakkar, Dweepna Garg, dan Pranay Jivrani	Penelitian ini melakukan analisis perbandingan antara model statistik tradisional ARIMA dengan tiga mode DL yaitu LSTM, BiLSTM, dan GRU untuk peramalan data deret waktu finansial. Model ini diuji pada empat dataset saham yang berbeda untuk memprediksi harga penutupan. Hasilnya menunjukkan bahwa ketiga model DL secara signifikan mengungguli model ARIMA. Di antara model-model <i>deep learning</i> , GRU menunjukkan kinerja terbaik dengan nilai <i>Root-Mean-Square-Error</i> (RMSE) rata-rata terendah yaitu 3.19. Secara keseluruhan, model GRU terbukti lebih baik 20,32% dari LSTM dan 9,8% lebih baik dari BiLSTM dalam hal akurasi prediksi.
3	IEEE Access, 9, 2021.	Detecting White Supremacist Hate Speech Using Domain Specific Word Embedding With Deep Learning and BERT [39],	Penelitian ini bertujuan untuk mendeteksi ujaran kebencian supremasi kulit putih di Twitter menggunakan teknik DL menggunakan model BiLSTM dengan <i>word embedding</i> spesifik domain yang diekstrak dari korpus

		Hind S. Alatawi, Areej M. Alhothali, dan Kawthar M. Moria.	supremasi kulit putih. model diuji pada dataset seimbang yang merupakan gabungan dari data Twitter dan forum Stormfront. Hasil penelitian menunjukkan bahwa model BiLSTM memberikan hasil dengan <i>domain-specific embedding</i> mencapai F1-score 0.75.
4	IEEE Sensors Journal, 24, 14, 2024.	A 3-D Attention-Enhanced Hybrid Neural Network for Turbofan Engine Remaining Life Prediction Using CNN and BiLSTM Models [40], You Keshun, Qiu Guangqi, dan Gu Yingkui.	Penelitian ini mengusulkan sebuah model jaringan saraf hibrida yang ditingkatkan dengan mekanisme <i>3-D attention</i> untuk prediksi sisa masa pakai (<i>Remaining Useful Life - RUL</i>) mesin turbofan. Model ini menggunakan <i>Convolutional Neural Network</i> (CNN) untuk mengekstrak fitur lokal dan <i>Bidirectional Long Short-Term Memory</i> (BiLSTM) untuk mempelajari fitur non-linear jangka panjang. Mekanisme <i>3-D attention</i> ditambahkan untuk menyeimbangkan bobot fitur dari sisi spasial dan kanal, yang juga meningkatkan interpretasi model. Model ini diuji pada dataset benchmark C-MAPSS. Hasilnya menunjukkan bahwa metode yang diusulkan memiliki kinerja yang paling unggul dibandingkan dengan metode prediksi RUL canggih lainnya, dengan rata-rata RMSE 14.85 dan Score 724.17.

5	2023 International Conference on Computer and Information Technology (ICCIT), 2023.	26th Next Word Prediction in Bangla Using Hybrid Approach[41], SM Nuruzzaman Nobel, Shirin Sultana, Md All Moon Tasir dan Md. Saifur Rahman.	Penelitian ini mengusulkan pendekatan hibrida untuk prediksi kata berikutnya dalam bahasa Bangla, yang menggabungkan struktur data Trie, <i>Convolutional Neural Network</i> (CNN), <i>Long Short-Term Memory</i> (LSTM), dan metodologi N-gram. Arsitektur hibrida ini dirancang untuk menangkap hubungan jangka panjang dan pola kontekstual yang kaya dalam bahasa Bangla. Model ini dievaluasi pada dataset komprehensif yang terdiri dari sekitar 50 ribu sampel bahasa Bangla dari berbagai sumber. Hasilnya menunjukkan bahwa model N-gram dengan urutan lebih tinggi sangat efektif. Secara spesifik, model 5-gram mencapai akurasi tertinggi sebesar 99,70%.
6	2023 International Conference on Electronics and Sustainable Communication Systems (ICESC), 2023.	4th A Machine Learning Approach to predict the Next Word in a Statement[42], Vishal Rathee, Sakshi Yede.	Penelitian ini melakukan studi perbandingan antara dua model <i>deep learning</i> untuk prediksi kata berikutnya (<i>Next-Word Prediction</i>), yaitu LSTM dan BiLSTM) Dataset yang digunakan adalah "medium-articles-dataset", yang datanya diperoleh melalui teknik <i>web scraping</i> . Kedua model diuji dengan menjalankan pelatihan selama 50 epoch untuk membandingkan akurasi dan kerugiannya. Hasil pengujian menunjukkan bahwa model BiLSTM secara signifikan lebih

			<p>baik daripada model LSTM. Model BiLSTM berhasil mencapai akurasi sekitar 85%, sementara model LSTM hanya mencapai akurasi sekitar 57%.</p>
7	<p>IEEE Access, 10, 2022.</p>	<p>Deep Learning-Based Approach for Inappropriate Content Detection and Classification of YouTube Videos[43], Kanwal Yousaf dan Tabassam Nawaz.</p>	<p>Studi ini mengusulkan arsitektur <i>deep learning</i> baru untuk mendeteksi dan mengklasifikasikan konten yang tidak pantas dalam video YouTube untuk anak-anak. Kerangka kerja yang diusulkan menggunakan model <i>Convolutional Neural Network (CNN) pre-trained</i> yang disebut EfficientNet-B7 untuk mengekstrak fitur video, yang kemudian dimasukkan ke dalam jaringan <i>Bidirectional Long Short-Term Memory (BiLSTM)</i>. Model ini dievaluasi pada dataset yang dianotasi secara manual yang terdiri dari 111.156 klip kartun dari YouTube. Hasil eksperimen menunjukkan bahwa arsitektur EfficientNet-BiLSTM dengan 128 unit tersembunyi memberikan kinerja terbaik. Model ini mencapai akurasi validasi sebesar 95,66% dan F1-score 0,9267 pada data pengujian.</p>

8	IEEE Access, 10, 2022.	Sentiment Analysis With Ensemble Hybrid Deep Learning Model[44], Kian Long Tan, Chin Poo Lee, Kian Ming Lim, dan Kalaiarasi Sonai Muthu Anbananthan.	Penelitian ini menyajikan sebuah model <i>ensemble hybrid deep learning</i> untuk analisis sentimen. Model yang diusulkan terdiri dari tiga model hibrida, yaitu RoBERTa-LSTM, RoBERTa-BiLSTM, dan RoBERTa-GRU, di mana prediksi dari ketiganya digabungkan menggunakan <i>averaging ensemble</i> dan <i>majority voting</i> . RoBERTa digunakan untuk menghasilkan <i>word embedding</i> yang representatif, sementara LSTM, BiLSTM, dan GRU bertugas menangkap dependensi jangka panjang. Model dievaluasi pada tiga dataset populer, termasuk IMDb dan Twitter US Airline Sentiment. Hasil eksperimen menunjukkan bahwa model <i>ensemble</i> ini mengungguli metode-metode canggih lainnya, dengan mencapai akurasi sebesar 94,9% pada dataset IMDb
9	IEEE Access, 9, 2021.	Batteries State of Health Estimation via Efficient Neural Networks With Multiple Channel Charging Profiles[45], Noman Khan, Fath U Min Ullah, Afnan, Amin Ullah, Mi Young Lee, dan Sung Wook Baik.	Menggunakan perbandingan antara metode ML dan DL untuk estimasi kapasitas dan <i>State of Health</i> (SOH) baterai dengan penggunaan <i>Multiple Channel Charging Profiles</i> (MCCPs), yang menggabungkan data voltase (V), arus (I), dan suhu (T), untuk meningkatkan akurasi estimasi. Berbagai model seperti SVR, <i>Adaptive Boosting</i> , Multi Layer Perceptron (MLP), CNN,

			LSTM, dan BiLSTM dievaluasi menggunakan dataset baterai dari NASA. Hasil eksperimen menunjukkan bahwa BiLSTM secara konsisten mengungguli semua pendekatan lain. Secara spesifik, model BiLSTM yang menggunakan MCCPs mencapai nilai <i>Mean Square Error</i> (MSE) terendah yaitu 0.0001 untuk Baterai #5.
10	IEEE Access, 2024	Advancing Fake News Detection: Hybrid Deep Learning With FastText and Explainable AI[15], Hashmi, Ehtesham; Yayilgan, Sule Yildirim; Yamin, Muhammad Mudassar; Ali, Subhan; dan Abomhara, Mohamed.	Model Bi-LSTM digunakan sebagai salah satu model <i>deep learning</i> yang dievaluasi, dengan input berupa <i>word embedding</i> dari FastText. Kinerjanya diukur pada tiga dataset berita palsu yang berbeda. Bi-LSTM menunjukkan kinerja yang sangat kompetitif. Pada dataset FakeNewsNet, model ini berhasil mencapai akurasi dan F1-score sebesar 0.97.
11	IEEE Access, 2022	A Deep Learning-Based Approach for Inappropriate Content Detection and Classification of YouTube Videos [46]. Yousaf, Kanwal dan Nawaz, Tabassam.	Model Bi-LSTM digunakan dalam arsitektur hybrid EfficientNet-BiLSTM. EfficientNet (sebuah CNN) mengekstrak fitur dari frame video, dan Bi-LSTM memproses urutan fitur tersebut untuk memahami konteks temporal dari video. Arsitektur hybrid yang menggunakan Bi-LSTM ini terbukti menjadi yang terbaik, dengan mencapai akurasi validasi 95,66% dan F1-score 0.9267. Ini menunjukkan bahwa Bi-LSTM sangat fleksibel dan kuat, tidak

			hanya untuk menganalisis teks, tetapi juga untuk memproses
--	--	--	---

Berbagai penelitian pada tabel 2.1 telah menunjukkan efektivitas dan keunggulan model Bidirectional Long Short-Term Memory (Bi-LSTM) dalam beragam tugas klasifikasi dan peramalan. Secara fundamental, Bi-LSTM terbukti secara signifikan lebih unggul dibandingkan LSTM satu arah, seperti yang ditunjukkan dalam tugas prediksi kata berikutnya di mana model Bi-LSTM mencapai akurasi sekitar 85%, jauh melampaui model LSTM yang hanya mencapai 57% [42].

Keunggulan Bi-LSTM juga terlihat pada domain di luar pemrosesan bahasa. Dalam peramalan data deret waktu untuk prediksi radiasi matahari, model Bi-LSTM menunjukkan kinerja terbaik di antara berbagai model deep learning lain yang diuji, dengan mencapai akurasi 96% [37]. Demikian pula, dalam estimasi kondisi baterai, Bi-LSTM ditemukan secara konsisten mengungguli semua pendekatan lain yang dievaluasi, termasuk CNN dan MLP [45].

Selain sebagai model mandiri, Bi-LSTM juga terbukti sangat kuat saat diintegrasikan dalam arsitektur hibrida. Pendekatan hybrid EfficientNet-BiLSTM untuk deteksi konten video berhasil mencapai akurasi validasi 95,66% [43]. Serupa dengan itu, model hybrid CNN-BiLSTM dengan 3-D attention untuk prediksi sisa masa pakai mesin juga terbukti memiliki kinerja paling unggul dibandingkan metode canggih lainnya [40]. Untuk konteks deteksi berita palsu yang relevan dengan penelitian ini, Bi-LSTM menunjukkan kinerja yang sangat kompetitif dengan mampu mencapai akurasi dan F1-score sebesar 0.97 pada dataset FakeNewsNet [15].

Secara kolektif, berbagai temuan ini menunjukkan bahwa Bi-LSTM merupakan pendekatan yang sangat baik, fleksibel, dan terbukti efektif untuk berbagai tugas klasifikasi dan peramalan sekuensial.