

Survei Literatur: Deteksi Berita Palsu Menggunakan Pendekatan *Deep Learning*

Rio Yunanto¹, Apriani Puti Purfini², Angga Prabuwisesa³

^{1,2}Program Studi Komputerisasi Akuntansi, FTIK UNIKOM, Bandung, Indonesia

³Program Studi Sistem Informasi, Universitas Indraprasta PGRI, Jakarta, Indonesia

e-mail: ¹rio.yunanto@email.unikom.ac.id, ²apriani.puti.purfini@email.unikom.ac.id, ³anggashela@gmail.com

Abstrak

Media sosial telah menjadi bagian keseharian yang tak terpisahkan di dalam kehidupan masyarakat modern untuk dapat memudahkan saling berinteraksi dan berkomunikasi. Tujuan dari penelitian survei ini, yaitu untuk mengulas dan membandingkan metode deep learning yang diimplementasikan pada kasus deteksi berita palsu dari beberapa penelitian sebelumnya, dan untuk mendapatkan gambaran korpus atau dataset yang digunakan oleh literatur sebelumnya. Penelitian ini juga untuk mengetahui dan memetakan pemanfaatan algoritma deep learning pada kasus deteksi berita palsu. Metode penelitian yang digunakan, yaitu melakukan survei literatur terhadap 12 literatur diperoleh dari website ScienceDirect dan IEEE Xplore. Kumpulan literatur yang disurvei tersebut diseleksi berdasarkan tahun terbit 2021 dengan topik penelitian deteksi berita palsu menggunakan pendekatan deep learning. Hasil survei literatur yang telah dilakukan merangkum bahwa, strategi mendeteksi berita palsu dapat dilakukan dengan empat pendekatan, yaitu berbasis isi/kontennya, berbasis gaya penulisan, pola penyebarannya, dan kredibilitas sumbernya. Hasil survei yang telah dilakukan juga memperlihatkan bahwa algoritma Convolutional Neural Network menjadi favorit pada penelitian-penelitian sebelumnya dengan muncul 6 kali di dalam kumpulan paper tersebut. Algoritma yang menjadi favorit berikutnya, yaitu Long Short Term Memory yang muncul pada 5 literatur dan Bidirectional LSTM yang muncul pada 4 literatur.

Kata kunci: artikel, bohong, klasifikasi, algoritma, media sosial

Abstract

Social media has become an inseparable part of everyday life in modern society to make it easier to interact and communicate with each other. The purpose of this study is to review and compare the deep learning methods implemented in the case of detecting fake news from several previous studies, and to get an overview of the corpus or dataset used by previous studies. This research is also to help researchers identify and map the use of deep learning algorithms in cases of detecting fake news. The research method is conducting a literature survey of 12 literatures obtained from the ScienceDirect and IEEE Xplore websites. The collection of literature that has been surveyed is selected based on the year published in 2021 with the topic of research on detection of fake news using a deep learning approach. The results of this study summarize that the strategy to detect fake news can be done with four approaches, based on the content, based on the writing style, based on the distribution pattern, and based on the credibility of the source. The results of this research also show that the Convolutional Neural Network algorithm is a favorite of researchers by appearing 6 times in the literature collection. The next favorite algorithm is Long Short Term Memory which appears in 5 literatures and Bidirectional LSTM which appears in 4 literatures.

Keywords: article, lie, classification, algorithm, social media

1. Pendahuluan

Kehidupan masyarakat modern hari ini telah sangat akrab dengan media sosial yang dapat memudahkan seseorang atau kelompok untuk saling berinteraksi dan berkomunikasi. Jika sebelumnya seseorang hanya dapat berkomunikasi dengan cara menghubungi orang tersebut melalui panggilan suara dan pesan teks, saat ini setiap aktifitas pengguna dapat diunggah ke media sosial sehingga orang lain dapat mengetahui dan mengikuti aktifitas pengguna tersebut tanpa harus menanyakan langsung melalui panggilan suara atau berkirim pesan. Penelitian terkait media sosial telah mampu membuat beberapa aplikasi yang mengintegrasikan konten pengguna media sosial dari berbagai wilayah tertentu untuk mengekstraksi informasi penting dalam waktu yang singkat. Salah satu contohnya berdasarkan kasus gempa bumi Jepang pada tahun 2009, dikembangkan algoritma yang menggabungkan pengguna Twitter sebagai sensor sosial untuk deteksi suatu peristiwa secara waktu nyata [1].

Media sosial sebagai bagian perkembangan teknologi yang tidak dapat dihindari, tentunya akan memiliki pengaruh positif dan negatif bagi penggunanya. Beberapa dampak negatif media sosial, antara lain; 1) privasi pengguna menjadi tidak ada batasan karena semua informasi diri dan aktivitas harian dapat setiap saat diunggah ke media sosial, 2) kejahatan pencurian data dan penipuan semakin meningkat karena pelaku kejahatan dapat berpura-pura menjadi seseorang tertentu menggunakan data pengguna di media sosial, 3) kasus perundungan digital dan potensi depresi terhadap anak-anak meningkat pesat, 4) penggiringan opini dengan menggunakan berita palsu meningkatkan potensi konflik horizontal terkait SARA (suku, ras, dan agama), dan lain-lain [2]. Media sosial secara tidak bijaksana mengakibatkan ketergantungan teknologi dan penggunaan yang berlebihan berdampak negatif pada organisasi. Secara khusus, terdapat tiga dimensi penggunaan media sosial yang berlebihan di tempat kerja, yaitu sosial yang berlebihan, hedonis, dan kognitif. Dimensi ini terkait dengan konflik penggunaan teknologi dan produktivitas kerja, yang pada gilirannya menurunkan kinerja kerja karyawan [3].

Berita palsu berdampak sangat buruk bagi penerima yang mudah percaya, bahkan dapat membunuh karakter manusia. Berita palsu dapat manipulasi individu atau kelompok dalam jangka waktu yang lama, tanpa disadari mental masyarakat akan terbentuk ke arah pemahaman yang salah. Penelitian sebelumnya melaporkan bahwa media sosial memiliki dampak yang signifikan terhadap penyebaran ketakutan dan kepanikan terkait wabah COVID-19 dan berpotensi munculnya pengaruh negatif terhadap kesehatan mental dan psikologis masyarakat. Korelasi statistik positif yang signifikan antara penggunaan media sosial dan penyebaran kepanikan terkait COVID-19 menunjukkan bahwa mayoritas remaja dan dewasa berusia 18-35 tahun menghadapi kecemasan psikologis [4].

Berbagai upaya terus dilakukan oleh pemerintah dan tokoh-tokoh masyarakat, bahkan komunitas sosial masyarakat juga melakukan edukasi pencegahan beredarnya berita palsu. Sebuah berita palsu yang muncul di media sosial bisa jadi karena disengaja atau tidak disengaja. Berita palsu yang cepat sekali viral dan tersebar luas, patut dicurigai sebagai bagian dari propaganda digital atau penggiringan opini publik. Diperlukan adanya suatu teknologi yang dapat melakukan deteksi dini berita palsu sehingga pihak-pihak terkait dapat melakukan penanggulangan dampak negatif yang terjadi.

Saat ini internet bukan lagi hal yang asing di Indonesia, internet bahkan sudah hadir di semua wilayah terdepan dan terluar di Indonesia meskipun dengan beberapa keterbatasan. Beberapa keterbatasan di wilayah terdepan dan terluar di Indonesia berupa sinyal seluler yang tidak stabil, dan ketersediaan listrik yang belum maksimal, tidak menyurutkan antusias pengguna internet di wilayah tersebut. Pengguna internet di Indonesia tercatat mencapai 171

juta penduduk, dan 95% dari pengguna internet tersebut memanfaatkan internet untuk beraktifitas di media sosial [5]. Aktifitas masyarakat yang cukup tinggi di media sosial dan layanan aplikasi pesan berakibat pada meningkatnya potensi terpapar berita palsu atau hoax. Survei yang dilakukan Mastel menunjukkan bahwa 62,80% penyebaran berita palsu atau hoax melalui jejaring media sosial dan aplikasi berkirim pesan [6].

Paparan yang telah diuraikan mengindikasikan bahwa tantangan dalam menangani penyebaran berita palsu atau berita bohong di media sosial yaitu dengan membangun deteksi dini berita palsu yang bekerja secara otomatis dan cepat. Mesin deteksi dini berita palsu tersebut harus mampu mengolah data dalam jumlah yang sangat besar dan diharapkan dapat mendeteksi penyebaran berita palsu secepat mungkin, agar dampak kerusakan sosial di masyarakat dapat dicegah sedini mungkin. Tujuan dari penelitian survei ini, yaitu untuk mengulas dan membandingkan metode *deep learning* yang diimplementasikan pada kasus deteksi berita palsu dari beberapa penelitian sebelumnya, dan untuk mendapatkan gambaran korpus atau *dataset* yang digunakan oleh penelitian sebelumnya. Hasil penelitian ini membantu untuk mengetahui dan memetakan pemanfaatan algoritma *deep learning* pada kasus deteksi berita palsu, dan dapat digunakan sebagai pondasi pengetahuan dalam menetapkan area penelitian yang telah matang pada topik deteksi berita palsu, sekaligus membuka sudut pandang penelitian di wilayah yang baru.

2. Kajian Pustaka

Berita palsu boleh jadi dapat dipahami sebagai berita bohong atau hoax, yang mana di dalam KBBI disebutkan bahwa hoax merupakan informasi yang direkayasa untuk menutupi informasi sebenarnya. Hoax juga dapat dipahami sebagai upaya pemutarbalikan fakta menggunakan informasi yang seolah-olah meyakinkan tetapi tidak dapat diverifikasi kebenarannya. Saat ini hoax tumbuh subur sebagai eksese negatif dari jejaring media sosial dan aplikasi pengiriman pesan. Salah satu contoh pemberitaan palsu yang paling umum adalah mengklaim sesuatu kejadian dengan suatu sebutan yang berbeda dengan kejadian yang sebenarnya untuk tujuan menghasut atau pembunuhan karakter pihak lain [7].

Menurut Zhou dan Zafarani deteksi berita palsu secara manual dapat dibagi menjadi; 1) pemeriksaan fakta berbasis ahli atau pakar, dan 2) pemeriksaan fakta berbasis bersumber data yang banyak. Pemeriksaan fakta manual berbasis ahli, yaitu mengandalkan ahli atau pakar di bidang pemeriksaan fakta menghasilkan data yang sangat kredibel, mudah dikelola, dan mengarah pada hasil yang sangat akurat, tetapi mahal. Deteksi berita palsu secara manual tidak sesuai dengan kondisi saat ini dimana volume informasi baru terus menerus muncul dengan cepat, terutama di media sosial. Salah satu solusi untuk mengatasi skalabilitas munculnya berita palsu, teknik pengecekan berita secara otomatis mulai dibutuhkan dan mengandalkan teknik *Natural Language Processing* (NLP) dan kecerdasan buatan. Penelitian terdahulu yang telah menemukan bahwa strategi yang dapat dilakukan untuk mendeteksi berita palsu terdiri dari empat perspektif, yaitu: 1) *knowledge-based*, 2) *style-based*, 3) *propagation-based*, dan 4) *source-based* [8].

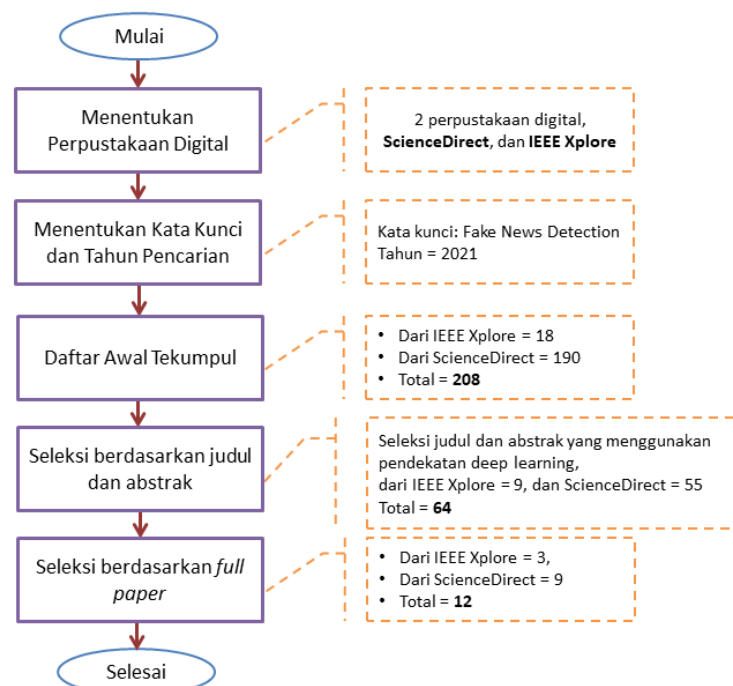
Machine learning adalah cabang dari bidang ilmu kecerdasan buatan dan menjadi salah satu cara yang dapat digunakan dalam kecerdasan buatan untuk mencapai hasil yang maksimal. *Machine learning* mengandalkan bekerja dengan kumpulan data kecil hingga besar dengan memeriksa dan membandingkan data tersebut untuk menemukan pola umum dan mengeksplorasi perbedaannya. Salah satu penerapan yang umum dari *machine learning* adalah prediksi hasil berdasarkan data yang ada. *Machine learning* mampu mempelajari pola dari kumpulan data yang ada, dan kemudian menerapkannya ke kumpulan data yang tidak

diketahui untuk memprediksi hasilnya. Teknik klasifikasi adalah teknik yang sering digunakan dalam *machine learning* untuk melakukan suatu proses prediksi hasil [9].

Deep learning dapat dipahami sebagai bentuk pembaruan dari *neural network multiple layer* yang merupakan bagian dari kecerdasan buatan dan *machine learning* yang dapat dimanfaatkan dalam tugas deteksi citra digital, pengenalan suara, dan lain-lain. Tahun 1950 tercatat bahwa *deep learning* mulai dikembangkan, akan tetapi bentuk penerapannya terlihat cukup memuaskan sejak tahun 1990. Algoritma pembelajaran yang digunakan sekarang pada tugas yang kompleks hampir sama seperti algoritma pembelajaran yang digunakan untuk menyelesaikan masalah permainan pada tahun 1980, meskipun model algoritma yang digunakan berubah menjadi training yang sederhana dari arsitektur *deep learning*. Hal yang penting pada pengembangan model yang sekarang, yaitu dukungan sumber daya perangkat keras yang dibutuhkan agar menjadi sukses [10]. *Deep learning* merupakan teknik modern terkini untuk pemrosesan gambar dan analisis data, dengan hasil yang menjanjikan dan potensinya sangat besar. Telah dilakukan survei terhadap 40 penelitian yang menggunakan teknik *deep learning*. Selain itu, telah dilakukan juga perbandingan *deep learning* dengan teknik populer lainnya, sehubungan dengan perbedaan kinerja teknik klasifikasi atau analisis regresi. Hasilnya menunjukkan bahwa *deep learning* memberikan akurasi tinggi, mengungguli teknik pemrosesan gambar yang umum digunakan [11].

3. Metode Penelitian

Studi literatur yang dilakukan adalah untuk mengetahui dan memetakan hasil studi terdahulu yang berkaitan dengan tema literatur tertentu. Studi literatur juga merupakan tahapan penting suatu penelitian karena dapat dijadikan sebagai pondasi pengetahuan, menetapkan area penelitian yang telah matang, dan membuka sudut pandang penelitian di wilayah yang baru. Studi literatur yang baik akan menghasilkan peta pengetahuan tentang suatu topik penelitian yang dapat memandu para peneliti menggali lebih dalam area yang belum matang [12]. Adapun metode penelitian yang digunakan seperti tampak pada Gambar 1.



Gambar 1. Diagram alur pencarian dan seleksi literatur

Data-data literatur pada penelitian ini dikumpulkan memanfaatkan pangkalan data penelitian ilmiah ScienceDirect dan IEEE Xplore dengan kata kunci “*Fake News Detection*” menggunakan pendekatan *deep learning*. Kumpulan literatur dengan topik deteksi berita palsu atau hoaks tersebut kemudian diseleksi secara bertahap, yaitu: 1) pendekatan yang digunakan hanya fokus pada model *deep learning*, 2) tahun terbit dari literatur yang didapatkan, diseleksi hanya tahun 2021 sebagaimana terlihat pada Gambar 1. Pencarian artikel menggunakan kata kunci dan tahun terbit artikel yang sudah ditentukan menghasilkan 208 artikel. Kemudian dilakukan seleksi judul dan abstrak berdasarkan kesesuaian topik yang diinginkan, menghasilkan 64 artikel. Kemudian dilakukan seleksi kembali dengan membaca *full paper* sehingga menghasilkan 12 artikel yang menjadi bahan survei literatur.

4. Hasil dan Pembahasan

Salah satu sisi gelap dari media sosial saat ini, yaitu menjadi sarana penyebaran informasi salah dan berita palsu. Informasi yang salah dapat mempengaruhi opini dan persepsi pembaca. Berita bohong bahkan dapat memanipulasi sentimen masyarakat untuk alasan politik atau ekonomi. Pengertian "berita palsu" atau "informasi salah" dapat dipahami sebagai informasi yang tidak akurat bahkan tidak jarang dibuat salah secara sengaja yang kemudian didistribusikan dengan maksud untuk menipu atau menggiring opini mereka yang membacanya.

Pada awal tahun 2020, sejumlah mitos mulai menjadi berita utama terkait dugaan pengobatan COVID-19, yang biasa disebut dengan virus corona. Banyak klaim palsu disajikan sebagai fakta, termasuk gagasan bahwa makan lebih banyak daging merah atau minyak kelapa dapat mengobati virus. Contoh misinformasi tersebut dapat membahayakan wacana publik dengan mempengaruhi individu untuk membuat keputusan yang salah terkait politik atau kesejahteraan pribadi. Secara sederhana misinformasi berarti salah informasi, yaitu konten informasi didalamnya memang salah, tetapi oleh pihak yang menyebarkan informasi tersebut percaya bahwa informasi itu benar. Penyebar informasi tak memiliki tendensi untuk membahayakan orang lain. Selain misinformasi ada juga disinformasi, yaitu pihak penyebar informasi mengetahui konten informasi didalamnya salah. Namun pihak penyebar tetap dengan sengaja menyebarkan informasi salah tersebut untuk menipu, mengancam, bahkan membahayakan pihak lain. Ada juga malinformasi, yaitu konten informasi yang didalamnya benar. Tetapi justru informasi tersebut digunakan untuk mengancam keberadaan pihak lain atau pihak dengan identitas tertentu. Malinformasi dapat bisa dikategorikan ke dalam hasutan kebencian [13].

Secara formal masih belum disepakati definisi universal untuk berita palsu, bahkan di dalam dunia jurnalisme. Pada dasarnya definisi yang jelas dan akurat dapat membantu untuk membedakan dan menganalisis berbagai jenis berita palsu, sehingga diperlukan beberapa konsep untuk menyajikan definisi yang luas dan sempit pada beberapa istilah berita palsu tersebut. Berdasarkan bagaimana istilah dan konsep berita palsu tersebut didefinisikan, maka dapat dibedakan satu dari yang lain berdasarkan tiga karakteristik utama, yaitu keaslian (mengandung pernyataan non-faktual atau tidak), niat (bertujuan untuk menyesatkan atau menghibur publik), dan konten informasi tersebut merupakan berita atau bukan. Rangkuman konsep-konsep berita palsu tersebut dapat dilihat pada Tabel 1., contohnya, disinformasi adalah informasi palsu (berita atau bukan berita) dengan niat jahat untuk menyesatkan publik [8].

Tabel 1. Perbandingan Konsep Berita Palsu [8]

Konsep	Keaslian	Maksud/Niat	Berita?
<i>Deceptive news</i>	Non-faktual	Menyesatkan	Ya
<i>False news</i>	Non-faktual	Tidak terdefinisi	Ya
<i>Satire news</i>	Tidak Utuh	Menghibur	Ya
<i>Disinformation</i>	Non-faktual	Menyesatkan	Tidak terdefinisi
<i>Misinformation</i>	Non-faktual	Tidak terdefinisi	Tidak terdefinisi
<i>Cherry-picking</i>	Biasanya faktual	Menyesatkan	Tidak terdefinisi
<i>Clickbait</i>	Tidak terdefinisi	Menyesatkan	Tidak terdefinisi
<i>Rumor</i>	Tidak terdefinisi	Tidak terdefinisi	Tidak terdefinisi

Tantangan penanggulangan berita palsu pada platform media sosial salah satunya, yaitu seringkali pendapat dan persepsi pribadi individu kemudian disalahpahami sebagai sebuah berita selayaknya berita dari sumber resmi. Bahkan tidak jarang terdapat pihak-pihak yang dengan sengaja menggunakan bot dan algoritma jahat di media sosial untuk melakukan propaganda secara paksa sehingga membuat pembaca percaya bahwa berita palsu itu sesuatu yang benar [14].

Berita palsu lebih jauh juga dapat mempengaruhi makro ekonomi suatu kawasan atau negara dimana tidak jarang berita palsu dapat mempengaruhi opini publik sehingga memberikan peluang bagi oknum pelaku kriminal keuangan untuk melakukan manipulasi pasar keuangan. Padahal berita keuangan yang valid merupakan sumber informasi utama yang sangat diperlukan bagi investor untuk melakukan analisis dan pengambilan keputusan investasi. Sehingga fokus deteksi berita palsu tidak hanya pada berita politik, melainkan juga berita keuangan perlu diperhatikan sebagai faktor penting dalam usaha deteksi berita palsu secara multidimensi [22].

Studi literatur yang telah dilakukan oleh penelitian sebelumnya juga melihat bahwa teknik *deep learning* telah berkembang dengan baik sehingga menjadi populer. Salah satu sebabnya, yaitu para peneliti tidak harus menghabiskan banyak waktu pada tahapan ekstraksi fitur untuk mendapatkan hasil prediksi yang andal menggunakan *deep learning*. Strategi pendekatan berbasis konten dan berbasis konteks juga menjadi salah satu cara yang efektif untuk melakukan klasifikasi berita palsu secara tepat [26]. Meskipun pada penelitian Reis dkk. pada tahun 2019 menunjukkan bahwa XGBoost atau eXtreme Gradient Boosting yang merupakan algoritma berbasis pohon telah menunjukkan performa yang sangat baik dalam melakukan deteksi berita palsu pada hitungan interval kepercayaan 95% untuk rata-rata AUC dan F1 dengan melakukan pembagian 5 kali lipat antara *dataset* pelatihan dan pengujian, dan diulang 10 kali dengan versi acak yang berbeda dari *dataset* asli (total 50 kali eksekusi) [27].

Survei literatur dari penelitian-penelitian yang telah dilakukan sebelumnya sebagaimana terlihat pada Tabel 2 bahwa telah diperoleh 12 literatur yang diterbitkan pada tahun 2021 dengan topik penelitian deteksi berita palsu menggunakan pendekatan *deep learning* yang diindeks oleh ScienceDirect, dan IEEE Xplore. Survei ini memperlihatkan bahwa terdapat 7 literatur yang menggunakan kumpulan data berbahasa Inggris sebagai objek penelitian, 1 literatur berbahasa Slavik, 1 literatur berbahasa India, 1 berbahasa Mandarin, 1 berbahasa Indonesia, dan 1 lagi menggunakan gabungan dua kumpulan data

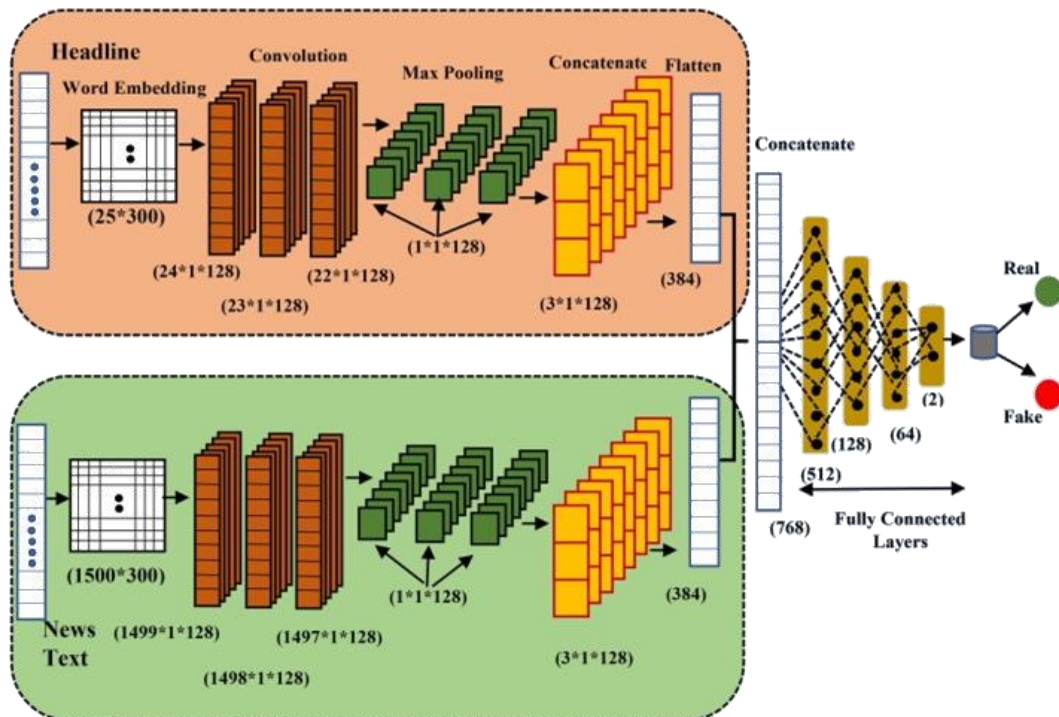
berbahasa Mandarin dan Inggris. Kondisi tersebut terjadi karena ketersediaan korpus data berita palsu berbahasa non Inggris yang sifatnya terbuka dan mudah diakses masih sangat terbatas. Pada penelitian Utami dkk.pada tahun 2020 telah melakukan implementasi algoritma Apriori dan Random Forest menggunakan korpus data berita berbahasa Indonesia dengan sumber data sebanyak 5.000 data berita pada periode 2016-2020 yang dikumpulkan dari turnbackhoax.id yang telah disusun klasifikasi kelas hoax dan non-hoax, dengan mengajukan permintaan [28]. Dataset berita palsu berbahasa Inggris seperti: LIAR Dataset, ISOT Dataset, dan FA-KES Dataset, juga sudah cukup populer dikalangan para peneliti sehingga memudahkan untuk menjadikannya sebagai tolak ukur tingkat keberhasilan algoritma atau kerangka kerja *deep learning* pada kasus deteksi berita palsu yang sedang dikembangkan.

Tabel 2. Perbandingan Tinjauan Pustaka Berita Palsu

No	Judul, Penulis, Tahun	Dataset / Objek Penelitian	Algoritma Deep Learning
1	<i>A Temporal Ensembling Based Semi-Supervised Convnet for The Detection of Fake News Articles.</i> , Meel & Vishwakarma, 2021 [14].	Kumpulan data berbahasa Inggris sebanyak 70.257 data dari 3 sumber, yaitu; Fake News Detection (Jruvika), Fake News Data, dan Fake News Sample (Pontes), yang tersedia di Kaggle.	<i>Convolutional Neural Network</i> (CNN)
2	<i>Attention-based C-BiLSTM for Fake News Detection.</i> , Trueman et al., 2021 [15].	Kumpulan data berbahasa Inggris sebanyak 12.836 data yang bersumber dari LIAR Dataset, dengan 6 label, yaitu; <i>true, half-true, mostly true, barely true, false, pants on fire.</i>	Kombinasi Bidirectional LSTM (Bi-LSTM) dan CNN di dalam AC-BiLSTM
3	<i>BerConvoNet: A Deep Learning Framework for Fake News Classification.</i> , M. Choudhary et al., 2021 [16].	Kumpulan data berbahasa Inggris sebanyak 54.924 data yang bersumber dari 4 sumber, yaitu; FakeNewsNet George McIntire Dataset, Repository, Gossipcop, Politifact.	CNN di dalam BerConvoNet yang menggunakan BERT sebagai encoder
4	<i>Comparison of Fake News Detection using Machine Learning and Deep Learning Techniques.</i> , Alameri & Mohd, 2021 [17].	Kumpulan data berbahasa Inggris sebanyak 18.113 data yang bersumber dari ISOT Dataset yang tersedia di Kaggle.	Long Short-Term Memory (LSTM)
5	<i>Convolutional Neural Network with Margin Loss for Fake News Detection.</i> , Goldani et al., 2021 [18].	Kumpulan data berbahasa Inggris sebanyak 57.734 data yang bersumber dari LIAR Dataset dan ISOT dataset.	CNN dengan margin loss
6	<i>Fake News Detection and Analysis using Multitask Learning with BiLSTM CapsNet Model.</i> , Sridhar & Sanagavarapu, 2021 [19].	Kumpulan data berbahasa Inggris sebanyak 20.554 data yang bersumber dari Fake News Dataset yang tersedia di Kaggle. Tersusun dalam 6 label kategori, yaitu; <i>not fake, may be misleading title, partial accurate, parody, misinformation, false/fake.</i>	Bidirectional LSTM (Bi-LSTM)

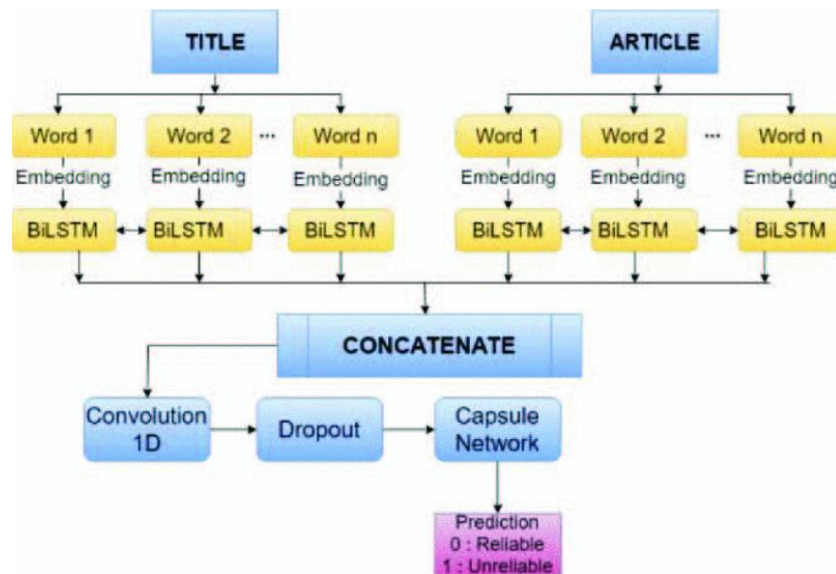
No	Judul, Penulis, Tahun	Dataset / Objek Penelitian	Algoritma Deep Learning
7	<i>Fake News Detection in Slovak Language using Deep Learning Techniques.</i> , Ivancova et al., 2021 [20].	Kumpulan data berbahasa Slavik sebanyak 2.278 data yang bersumber dari 9 website, berisi potongan berita.	Kombinasi CNN dan LSTM
8	<i>Fake News Detection: A Hybrid CNN-RNN Based Deep Learning Approach.</i> , Nasir et al., 2021 [21].	Kumpulan data berbahasa Inggris sebanyak 45.804 data yang bersumber dari ISOT Dataset dan FA-KES Dataset.	Kombinasi CNN dan RNN
9	<i>Financial Fake News Detection with Multi fact CNN-LSTM Model.</i> , Zhi et al., 2021 [22].	Kumpulan data berbahasa Mandarin sebanyak 8.000 data berlabel, memiliki judul, konten, sumber, dan komentar.	Kombinasi CNN dan LSTM
10	<i>Hoax Analyzer for Indonesian News Using Deep Learning Models.</i> , Nayoga et al., 2021 [23].	Kumpulan data berbahasa Indonesia sebanyak 1.000 data yang bersumber dari data.mendeley.com/datasets , github.com/pierobeat/ , dan kompas.com .	DNN, LSTM, Bi-LSTM, GRU, Bi-GRU, dan 1D-CNN
11	<i>Improving Fake News Detection with Domain-Adversarial and Graph-attention Neural Network.</i> , Yuan et al., 2021 [24].	Kumpulan data berbahasa Mandarin dan berbahasa Inggris sebanyak 15.400 yang bersumber dari Twitter dan Weibo, berisi gambar dan teks, diberi label asli atau palsu.	Kombinasi Bi-LSTM dan Graph Neural Network di dalam DAGANN (GNN)
12	<i>Multiple Features Based Approach for Automatic Fake News Detection on Social Networks using Deep Learning.</i> , Sahoo & Gupta, 2021 [25].	Kumpulan data berbahasa India sebanyak 42.277.247 yang bersumber dari Facebook, dengan rincian, yaitu; 5.026 data <i>profiles</i> , 15.328 data berita, dan 42.256.893 data <i>post</i> .	<i>Long Short Term Memory</i> (LSTM)

Algoritma *deep learning* yang digunakan pada 12 literatur tersebut juga memperlihatkan bahwa algoritma *Convolutional Neural Network* (CNN) menjadi favorit dengan muncul 6 kali di dalam kumpulan paper tersebut. Algoritma yang menjadi favorit berikutnya, yaitu *Long Short Term Memory* (LSTM) yang muncul pada 5 literatur dan *Bidirectional LSTM* (Bi-LSTM) yang muncul pada empat literatur. Algoritma CNN tampak cukup mudah untuk diterapkan pada kasus-kasus yang berkaitan dengan NLP khususnya dalam hal ini deteksi berita palsu. CNN juga memiliki kinerja yang bagus dan tingkat akurasi yang tinggi pada hasilnya. Kesuksesan CNN ternyata masih menyisakan kekurangan, yaitu hanya dapat dioperasikan pada data terstruktur reguler (Euclidean), sementara kenyataannya banyak juga data di dunia nyata yang di mana jumlah koneksi *node*-nya bervariasi, tidak berurutan, tidak teratur, atau non-Euclidean. Sebagaimana pada penelitian Benamira pada tahun 2019 yang melakukan klasifikasi atas kesamaan *graph* menggunakan 2 metode, yaitu: *Graph Convolutional Networks* (GCN) dan *Attention Graph Neural Network* (AGNN) untuk mendeteksi berita palsu pada sumber *dataset* dari “Horne 2017 Fake News Data” yang terdiri dari 150 data berita yang berlabel, 75 palsu dan 75 asli [29].



Gambar 2. Diagram arsitektur CNN untuk deteksi berita palsu [14]

Konsep kerja CNN yang diterapkan untuk kasus deteksi berita palsu sebagaimana pada tampilan Gambar 2, terlihat bahwa sampel data masukkan dibedakan menjadi 2 bentuk yaitu judul berita dan teks konten berita. Penggunaan Word2Vec *word embedding* yaitu untuk mewakili setiap kata inputan ke ruang vektor 300 dimensi. Word2Vec merupakan metode yang dapat menghasilkan teknik pemrosesan bahasa alami yang digunakan untuk mewakili kata-kata dalam bentuk vektor numerik dalam ruang geometris dari makna semantiknya, di mana jarak antara dua vektor akan menggambarkan hubungan semantik antara dua kata tersebut berkorelasi positif dengan kesamaan semantik kata-kata tersebut. Batas maksimum tensor input pada judul yaitu 25 kata sedangkan pada teks konten berita yaitu 1500 kata, yang kemudian dilanjutkan pada tahap proses kernel konvolusi, *max polling*, *concatenate*, dan *flatten*. Gabungan *reshape* fitur menjadi sebuah vector agar bisa digunakan sebagai input pada tahap proses *fully-connected layer* [14]. *Fully-connected layer* yaitu suatu susunan neuron-neuron yang semuanya saling berhubungan dengan neuron lain pada lapisan sebelumnya. Setiap aktivasi dari lapisan sebelumnya perlu diubah menjadi data satu dimensi sebelum dapat terhubung ke semua neuron, kemudian layer terakhir (*loss layer*) pada CNN yang bertugas mengukur penyimpangan nilai prediksi dengan nilai aktual (target) lalu menghasilkan klasifikasi berupa label berita palsu atau asli [30].



Gambar 3. Diagram arsitektur Bi-LSTM untuk deteksi berita palsu [19]

Arsitektur BiLSTM untuk deteksi berita palsu sebagaimana pada Gambar 3., terlihat bahwa model yang dirancang mengambil data berita seperti konsep kerja CNN yaitu dengan dibedakan menjadi 2 bentuk yaitu judul berita dan teks konten berita. Kata-kata dari judul berita dan teks konten berita divektorkan ke dalam *embedding matrix* dan dimasukkan ke dalam lapisan Bi-LSTM. Judul dan isi berita kemudian masing-masing dimodelkan menggunakan unit Bi-LSTM dan hasilnya kemudian digabungkan kembali. Vektor gabungan tersebut kemudian diproses lagi menggunakan lapisan konvolusi dan kemudian dimasukkan ke dalam *capsule network* untuk diprediksi apakah data berita tersebut asli atau palsu [19].

Jika dimungkinkan untuk mengelompokkan algoritma LSTM dan Bi-LSTM sebagai satu kelompok yang sama, tentunya kelompok algoritma LSTM ini menjadi lebih populer dibandingkan algoritma CNN berdasarkan kumpulan literatur yang telah dikumpulkan. LSTM merupakan algoritma yang memiliki keunggulan dibandingkan RNN (Recurrent Neural Network) karena memiliki dua jenis ingatan yaitu memori jangka panjang dan jangka pendek yang memungkinkan untuk menyimpan informasi terhadap pola-pola data. Perbandingan CNN dengan LSTM dalam melakukan deteksi berita palsu berbahasa Indonesia telah mampu menghasilkan nilai *accuracy*, *precision*, dan *recall* lebih dari 0,88 untuk CNN, dan lebih dari 0,84 untuk LSTM yang diimplementasikan pada kumpulan data yang dimiliki Masyarakat Anti Hoax Indonesia (MAFINDO) dan Forum Anti Fitnah Hasut dan Hoax (FAFHH) [31]. Keunggulan Bi-LSTM sedangkan dua lapisan jaringan saraf yang dapat menangani data sekuensial untuk meningkatkan akurasi prediksi dengan bergerak maju dan mundur. Beberapa catatan yang menarik salah satunya bahwa di balik keunggulan Bi-LSTM tersebut, masih terdapat beberapa permasalahan, salah satunya yaitu: arsitektur algoritma Bi-LSTM yang kompleks menjadi potensi beban komputasi yang tinggi ketika diterapkan pada kasus berskala besar.

5. Kesimpulan

Berdasarkan survei literatur pada 12 artikel yang terpilih, secara garis besar bahwa algoritma CNN dan Bi-LSTM menjadi favorit para peneliti saat ini untuk diimplementasikan pada kasus pendeteksian berita palsu karena keunggulan Bi-LSTM yang memiliki dua

lapisan jaringan saraf yang dapat menangani data sekuensial untuk meningkatkan akurasi prediksi dengan bergerak maju dan mundur. Sedangkan menurut Zhou dan Zafarani bahwa deteksi berita palsu secara manual yang melibatkan pemeriksaan fakta berbasis ahli/pakar dapat menghasilkan data yang sangat kredibel, tetapi berbiaya mahal dan membutuhkan waktu cukup lama. Deteksi berita palsu secara manual tidak sesuai dengan volume informasi baru yang terus menerus muncul dengan cepat, terutama di media sosial. Tingkat masivitas berita palsu yang sangat tinggi hanya mungkin diatasi dengan teknik deteksi berita palsu secara otomatis dengan mengandalkan pendekatan NLP dan kecerdasan buatan. Strategi mendeteksi berita palsu dapat dilakukan dengan empat pendekatan, yaitu; berbasis isi/kontennya, berbasis gaya penulisannya, berbasis pola penyebarannya, dan berbasis kredibilitas sumbernya.

Ucapan Terima Kasih

Penulis mengucapkan terima kasih kepada Fakultas Teknik dan Ilmu Komputer, Universitas Komputer Indonesia, atas dukungan teknisnya. Penelitian yang disajikan dalam makalah ini telah dilakukan di Laboratorium Sistem Informasi Akuntansi, Prodi Komputerisasi Akuntansi.

Daftar Pustaka

- [1] C. Reuter and M. A. Kaufhold, "Fifteen years of social media in emergencies: A retrospective review and future directions for crisis Informatics," *J. Contingencies Cris. Manag.*, vol. 26, no. 1, 2018, doi: 10.1111/1468-5973.12196.
- [2] N. Hidayat, N. Qalby, S. S. Alaydrus, A. Darmayanti, and A. P. Salsabila, "Pengaruh Media Sosial Terhadap Penyebaran Hoax Oleh Digital Native," Makassar, 2019.
- [3] X. Cao and L. Yu, "Exploring The Influence Of Excessive Social Media Use At Work: A Three-Dimension Usage Perspective," *Int. J. Inf. Manage.*, vol. 46, no. 1, pp. 83–92, 2019.
- [4] A. R. Ahmad and H. R. Murad, "The Impact of Social Media on Panic During The COVID-19 Pandemic in Iraqi Kurdistan: Online Questionnaire Study," *J. Med. Internet Res.*, vol. 22, no. 5, 2020.
- [5] M. R. Ramadhani and A. R. I. Pratama, "Analisis Kesadaran Cyber Security Pada Pengguna Media Sosial Di Indonesia," *AUTOMATA*, vol. 1, no. 2, 2020, [Online]. Available: <https://journal.uui.ac.id/AUTOMATA/article/download/15426/10219>.
- [6] S. Gerintya, "Hoaks dan Bahaya Rendahnya Kepercayaan Terhadap Media," 2018. <https://tirto.id/hoaks-dan-bahaya-rendahnya-kepercayaan-terhadap-media-cKAX> (accessed Jul. 28, 2021).
- [7] E. Lararenjana, "Mengenal Arti Hoax Atau Berita Bohong, Ketahui Jenis dan Ciri-Cirinya," 2020. <https://www.merdeka.com/jatim/mengenal-arti-hoax-atau-berita-bohong-dan-cara-tepat-menyikapinya-klN.html?page> (accessed Jul. 28, 2021).
- [8] X. Zhou and R. Zafarani, "A Survey of Fake News: Fundamental Theories, Detection Methods, and Opportunities," *ACM Comput. Surv.*, vol. 53, no. 5, pp. 1–40, Oct. 2020, doi: 10.1145/3395046.
- [9] C. B. C. Latha and S. C. Jeeva, "Improving The Accuracy Of Prediction Of Heart Disease Risk Based On Ensemble Classification Techniques," *Informatics Med. Unlocked*, vol. 16, 2019, doi: 10.1016/j.imu.2019.100203.
- [10] A. Yanuar, "Pengenalan Deep Learning," *ugm.ac.id*, 2018. <https://machinelearning.mipa.ugm.ac.id/2018/06/10/pengenalan-deep-learning> (accessed Jul. 28, 2021).

- [11] A. Kamilaris and F. X. Prenafeta-Boldú, "Deep Learning In Agriculture: A Survey," *Computers and Electronics in Agriculture*, vol. 147. 2018, doi: 10.1016/j.compag.2018.02.016.
- [12] C. Fisch and J. Block, "Six tips for your (systematic) literature review in business and management research," *Manag. Rev. Q.*, vol. 68, no. 2, pp. 103–106, 2018, doi: 10.1007/s11301-018-0142-x.
- [13] R. Asyik, "Inilah Beda Misinformasi, Disinformasi, dan Malinformasi," *ayobandung.com*, 2019. <https://ayobandung.com/read/2019/02/01/44283/inilah-beda-misinformasi-disinformasi-dan-malinformasi>.
- [14] P. Meel and D. K. Vishwakarma, "A Temporal Ensembling Based Semi-Supervised Convnet for The Detection of Fake News Articles," *Expert Syst. Appl.*, vol. 177, p. 115002, Sep. 2021, doi: 10.1016/j.eswa.2021.115002.
- [15] T. E. Trueman, A. K. J., N. P., and V. J., "Attention-based C-BiLSTM for Fake News Detection," *Appl. Soft Comput.*, vol. 110, p. 107600, Oct. 2021, doi: 10.1016/j.asoc.2021.107600.
- [16] M. Choudhary, S. S. Chouhan, E. S. Pilli, and S. K. Vipparthi, "BerConvoNet: A Deep Learning Framework for Fake News Classification," *Appl. Soft Comput.*, vol. 110, p. 107614, Oct. 2021, doi: 10.1016/j.asoc.2021.107614.
- [17] S. A. Alameri and M. Mohd, "Comparison of Fake News Detection using Machine Learning and Deep Learning Techniques," in *2021 3rd International Cyber Resilience Conference (CRC)*, Jan. 2021, pp. 1–6, doi: 10.1109/CRC50527.2021.9392458.
- [18] M. H. Goldani, R. Safabakhsh, and S. Momtazi, "Convolutional Neural Network with Margin Loss for Fake News Detection," *Inf. Process. Manag.*, vol. 58, no. 1, p. 102418, Jan. 2021, doi: 10.1016/j.ipm.2020.102418.
- [19] S. Sridhar and S. Sanagavarapu, "Fake News Detection and Analysis using Multitask Learning with BiLSTM CapsNet Model," in *2021 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence)*, Jan. 2021, pp. 905–911, doi: 10.1109/Confluence51648.2021.9377080.
- [20] K. Ivancova, M. Sarnovski, and V. Maslej-Krcsnakova, "Fake News Detection in Slovak Language using Deep Learning Techniques," in *2021 IEEE 19th World Symposium on Applied Machine Intelligence and Informatics (SAMII)*, Jan. 2021, pp. 000255–000260, doi: 10.1109/SAMII50585.2021.9378650.
- [21] J. A. Nasir, O. S. Khan, and I. Varlamis, "Fake News Detection: A Hybrid CNN-RNN Based Deep Learning Approach," *Int. J. Inf. Manag. Data Insights*, vol. 1, no. 1, p. 100007, Apr. 2021, doi: 10.1016/J.JJIMEI.2020.100007.
- [22] X. Zhi *et al.*, "Financial Fake News Detection with Multi fact CNN-LSTM Model," in *2021 IEEE 4th International Conference on Electronics Technology (ICET)*, May 2021, pp. 1338–1341, doi: 10.1109/ICET51757.2021.9450924.
- [23] B. P. Nayoga, R. Adipradana, R. Suryadi, and D. Suhartono, "Hoax Analyzer for Indonesian News Using Deep Learning Models," *Procedia Comput. Sci.*, vol. 179, pp. 704–712, 2021, doi: 10.1016/j.procs.2021.01.059.
- [24] H. Yuan, J. Zheng, Q. Ye, Y. Qian, and Y. Zhang, "Improving Fake News Detection with Domain-Adversarial and Graph-attention Neural Network," *Decis. Support Syst.*, p. 113633, Jul. 2021, doi: 10.1016/j.dss.2021.113633.
- [25] S. R. Sahoo and B. B. Gupta, "Multiple Features Based Approach for Automatic Fake News Detection on Social Networks using Deep Learning," *Appl. Soft Comput.*, vol. 100, 2021, doi: 10.1016/j.asoc.2020.106983.
- [26] A. Bondielli and F. Marcelloni, "A survey on fake news and rumour detection

- techniques,” *Inf. Sci. (Ny)*, vol. 497, 2019, doi: 10.1016/j.ins.2019.05.035.
- [27] J. C. S. Reis, A. Correia, F. Murai, A. Veloso, F. Benevenuto, and E. Cambria, “Supervised Learning for Fake News Detection,” *IEEE Intell. Syst.*, vol. 34, no. 2, pp. 76–81, Mar. 2019, doi: 10.1109/MIS.2019.2899143.
- [28] M. P. Utami, O. D. Nurhayati, and B. Warsito, “Hoax Information Detection System Using Apriori Algorithm and Random Forest Algorithm in Twitter,” 2020, doi: 10.1109/ICIDM51048.2020.9339648.
- [29] A. Benamira, B. Devillers, E. Lesot, A. K. Ray, M. Saadi, and F. D. Malliaros, “Semi-supervised learning and graph neural networks for fake news detection,” 2019, doi: 10.1145/3341161.3342958.
- [30] H. K. Farid, E. B. Setiawan, and I. Kurniawan, “Implementation Information Gain Feature Selection for Hoax News Detection on Twitter using Convolutional Neural Network (CNN),” *Indones. J. Comput.*, vol. 5, no. 3, pp. 23–36, 2020.
- [31] A. A. Kurniawan and M. Mustikasari, “Implementasi Deep Learning Menggunakan Metode CNN dan LSTM untuk Menentukan Berita Palsu dalam Bahasa Indonesia,” *J. Inform. Univ. Pamulang*, vol. 5, no. 4, pp. 544–552, 2021.