© Jurnal Nasional Teknik Elektro dan Teknologi Informasi
Karya ini berada di bawah Lisensi Creative Commons Atribusi-BerbagiSerupa 4.0 Internasional
DOI: 10.22146/jnteti.v14i2.19050

Deteksi Berita Hoaks Berbahasa Indonesia Menggunakan *One-Dimensional Convolutional Neural Network*

Muhammad Zuama Al Amin[1], Muhammad 'Ariful Furqon[2], Dwi Wijonarko[1]

[1] Program Studi Teknologi Informasi, Fakultas Ilmu Komputer, Universitas Jember, Jember, Jawa Timur 68121, Indonesia
[2] Program Studi Informatika, Fakultas Ilmu Komputer, Universitas Jember, Jember, Jawa Timur 68121, Indonesia

[Diserahkan: 25 Januari 2025, Direvisi: 14 April 2025, Diterima: 28 Mei 2025]
Penulis Korespondensi: Muhammad 'Ariful Furqon (email: ariful.furqon@unej.ac.id)

INTISARI — Kemajuan teknologi informasi yang pesat telah memfasilitasi penyebaran informasi secara global, tetapi juga meningkatkan prevalensi berita hoaks, khususnya di Indonesia. Berita hoaks memiliki potensi untuk menciptakan disinformasi yang dapat memengaruhi opini publik, stabilitas sosial, dan keamanan. Oleh karena itu, diperlukan solusi berbasis teknologi yang efektif untuk mendeteksi dan mengidentifikasi hoaks atau berita palsu. Penelitian ini bertujuan untuk mengembangkan dan mengoptimalkan model *one-dimensional convolutional neural network* (1D-CNN) guna mendeteksi berita hoaks dengan tingkat akurasi yang tinggi. *Dataset* yang digunakan terdiri atas 12.151 artikel, yang mencakup 5.276 berita valid dan 6.875 berita hoaks, yang diperoleh dari sumber tepercaya serta platform antihoaks. Tahapan prapemrosesan teks meliputi pembersihan data, *case folding*, penghapusan tanda baca, penghapusan angka, dan penghapusan *stopword*. Data tekstual selanjutnya diproses melalui tahapan tokenisasi dan *padding* untuk persiapan pelatihan model. Arsitektur 1D-CNN yang diusulkan mengintegrasikan lapisan-lapisan *embedding*, *conv1d*, *batch normalization*, *globalmaxpooling1d*, *dense*, dan *dropout*, untuk meningkatkan kemampuan generalisasi serta mengurangi risiko *overfitting*. Model dilatih menggunakan *optimizer* Adam dan evaluasi kinerja menggunakan 10-*fold cross-validation*. Hasil eksperimen menunjukkan bahwa model menghasilkan rata-rata akurasi 97,74%, presisi 97,75%, *recall* 97,74%, dan *F*1-*score* 97,73%. Model yang dikembangkan mampu mengungguli metode sebelumnya, seperti kombinasi *convolutional neural network- bidirectional long short – term memory neural network* (CNN-BiLSTM), *gated recurrent unit* (GRU), dan metode konvensional seperti *naïve Bayes* atau *support vector machine* (SVM), baik dari segi akurasi maupun efisiensi waktu pelatihan. Penelitian ini menunjukkan bahwa model memiliki kemampuan yang andal dalam mengidentifikasi berita hoaks, baik dari segi tingkat deteksi yang akurat maupun konsistensi kinerja.

KATA KUNCI — 1D-CNN, Deteksi Berita Hoaks, Klasifikasi Teks, Pemelajaran Mendalam.

I. PENDAHULUAN

Perkembangan teknologi mempermudah akses informasi dan penyebarannya, terutama melalui internet. Platform digital seperti media sosial dan situs berita mempercepat serta memperluas interaksi antarindividu. Namun, kemudahan ini juga membuka peluang besar bagi penyebaran informasi hoaks, yang sering kali sulit dikenali. Menurut Kamus Besar Bahasa Indonesia (KBBI), hoaks adalah informasi palsu yang sengaja disebarkan untuk menipu atau menyesatkan [1]. Hoaks sendiri sulit dikenali karena derasnya arus informasi yang menyulitkan verifikasi. Media utama penyebar hoaks meliputi blog, situs berita, dan media sosial [2].

Data dari Kementerian Komunikasi dan Informatika menunjukkan bahwa ada lebih dari 10.000 kasus hoaks dilaporkan di Indonesia dalam lima tahun terakhir [3]. Fenomena ini juga diperburuk oleh rendahnya tingkat literasi informasi masyarakat Indonesia. Menurut survei PISA pada tahun 2022, peringkat literasi Indonesia berada di posisi 71 dari 81 negara [4]. Kondisi ini mengakibatkan masyarakat cenderung menerima informasi secara pasif tanpa melakukan verifikasi, sehingga memperbesar potensi penyebaran hoaks.

Untuk menghadapi tantangan ini, diperlukan solusi yang dapat mendeteksi berita hoaks yang efektif guna membantu masyarakat dalam mengidentifikasi informasi valid dan hoaks [1]. Sistem seperti ini dapat mengurangi penyebaran hoaks dengan menyediakan alat yang praktis dan akurat untuk memverifikasi informasi. Berbagai pendekatan berbasis kecerdasan buatan, pemelajaran mesin, dan pemelajaran mendalam telah dikembangkan untuk mengatasi masalah ini. Namun, sebelum membangun sistem deteksi berbasis pemelajaran mendalam atau pemelajaran mesin, diperlukan pemilihan model yang sesuai untuk memastikan sistem dapat diimplementasikan secara efektif [1].

Pendekatan utama dalam mendeteksi berita hoaks adalah pemelajaran terbimbing (*supervised learning*). Model dilatih dengan data berlabel untuk mengenali pola pada teks [5]. Dalam konteks ini, deteksi berita hoaks sering diformulasikan sebagai masalah klasifikasi atau regresi, meskipun pendekatan klasifikasi lebih banyak diterapkan [5]. Di antara berbagai metode klasifikasi, model pemelajaran mendalam lebih menonjol dibandingkan metode pemelajaran mesin biasa karena kemampuannya untuk mengenali pola-pola kompleks dalam data teks.

Salah satu model pemelajaran mendalam yang telah terbukti memiliki tingkat akurasi tinggi dalam mendeteksi berita hoaks adalah *one-dimensional convolutional neural network* (1D-CNN). Model ini memproses teks sebagai vektor 1D, yang memungkinkan pengenalan pola-pola penting dalam data dengan efisiensi tinggi. Dibandingkan dengan metode tradisional seperti *naïve Bayes* dan *support vector machine* (SVM) [1], 1D-CNN menunjukkan keunggulan signifikan baik dalam hal akurasi maupun kemampuan generalisasi. Selain itu,

karena model 1D-CNN merupakan model yang memiliki arsitektur sederhana, model ini sangat rentan terhadap *overfitting* bila diimplementasikan dalam *dataset* yang besar.

Penelitian sebelumnya menunjukkan bahwa model 1D-CNN sangat efektif dalam mendeteksi berita hoaks berbahasa Indonesia. Model ini dinilai fleksibel karena dapat dikombinasikan dengan teknik lain dan mampu mencapai tingkat akurasi yang tinggi. Referensi [1] menunjukkan bahwa 1D-CNN yang dikombinasikan dengan teknik *batch normalization* mampu mendeteksi berita hoaks berbahasa Indonesia dengan akurasi mencapai 97,9%. Selain itu, 1D-CNN berhasil mendeteksi hoaks dalam berita berbahasa Indonesia dengan akurasi sebesar 99% [3]. Dalam [6], model 1D-CNN mencapai akurasi 96,51% dalam mendeteksi berita hoaks di media sosial. Keunggulan ini menegaskan bahwa 1D-CNN merupakan pendekatan yang sangat menjanjikan untuk diterapkan dalam skenario praktis, terutama di tengah meningkatnya kebutuhan akan solusi cerdas demi menangkal penyebaran informasi hoaks.

Penelitian ini bertujuan untuk mengoptimalkan model 1D-CNN dengan menerapkan *batch normalization* dan penambahan lapisan (*layer*) *dropout* guna meningkatkan generalisasi model sekaligus mengurangi risiko *overfitting* [7]. Tujuan utama dari penelitian ini adalah mengevaluasi kinerja model 1D-CNN dalam mendeteksi berita hoaks berbasis data teks, khususnya dalam konteks berita berbahasa Indonesia, serta mendapatkan kinerja optimal dari model 1D-CNN. Penelitian ini menawarkan pendekatan inovatif dengan mengoptimalkan arsitektur 1D-CNN melalui kombinasi strategi lapisan *dropout* dan *batch normalization*, yang dirancang untuk mengatasi tantangan *overfitting* pada implementasi model pemelajaran mendalam terhadap *dataset* berskala besar. Pendekatan ini diharapkan dapat meningkatkan efisiensi dan akurasi model dalam mendeteksi pola-pola linguistik dan semantik pada data berita hoaks.

Kontribusi utama penelitian ini terletak pada optimalisasi model deteksi hoaks berbasis pemelajaran mendalam yang dilatih menggunakan *dataset* berita berbahasa Indonesia. Optimalisasi dilakukan dengan mengimplementasikan arsitektur model 1D-CNN yang dilengkapi dengan *batch normalization* dan penambahan lapisan *dropout*, yang bertujuan untuk meningkatkan kinerja model dalam membedakan berita hoaks dari berita yang valid. Penelitian ini juga mencakup evaluasi menyeluruh terhadap kinerja model pada data teks dalam konteks lokal, yang memberikan wawasan baru mengenai pengembangan teknologi deteksi informasi palsu yang lebih spesifik dan relevan di Indonesia.

Manfaat utama penelitian ini adalah menyediakan model pemelajaran mendalam sebagai solusi praktis untuk mendeteksi berita hoaks secara cepat dan akurat. Model yang dikembangkan diharapkan tidak hanya meningkatkan efektivitas deteksi informasi palsu, tetapi juga berkontribusi pada pengembangan dan optimalisasi teknologi pemelajaran mendalam dalam deteksi berita hoaks berbahasa Indonesia.

II. BERITA HOAKS

Menurut Kamus Besar Bahasa Indonesia (KBBI), berita palsu atau hoaks didefinisikan sebagai informasi yang tidak benar, tetapi sengaja dibuat dan disebarkan seolah-olah merupakan kebenaran. Informasi ini bertujuan untuk menipu, menyesatkan, atau memanipulasi penerima informasi agar percaya pada sesuatu yang sebenarnya tidak terjadi. Hoaks sering kali disebarkan melalui berbagai media, terutama di era digital saat ini, yang penyebaran informasi dapat dilakukan dengan cepat dan tanpa verifikasi yang memadai [2].

Salah satu karakteristik berita yang dapat membedakan bahwa suatu berita merupakan hoaks atau bukan adalah gaya penulisan yang provokatif dan berlebihan. Hoaks cenderung menggunakan kalimat bombastis yang dirancang untuk memancing emosi pembaca [8], seperti kemarahan atau ketakutan, berbeda dengan berita sahih atau valid yang memiliki gaya penulisan lebih formal dan objektif. Perbedaan dalam tata kalimat dan gaya penulisan ini menjadi salah satu indikator penting untuk memverifikasi keaslian suatu berita [2].

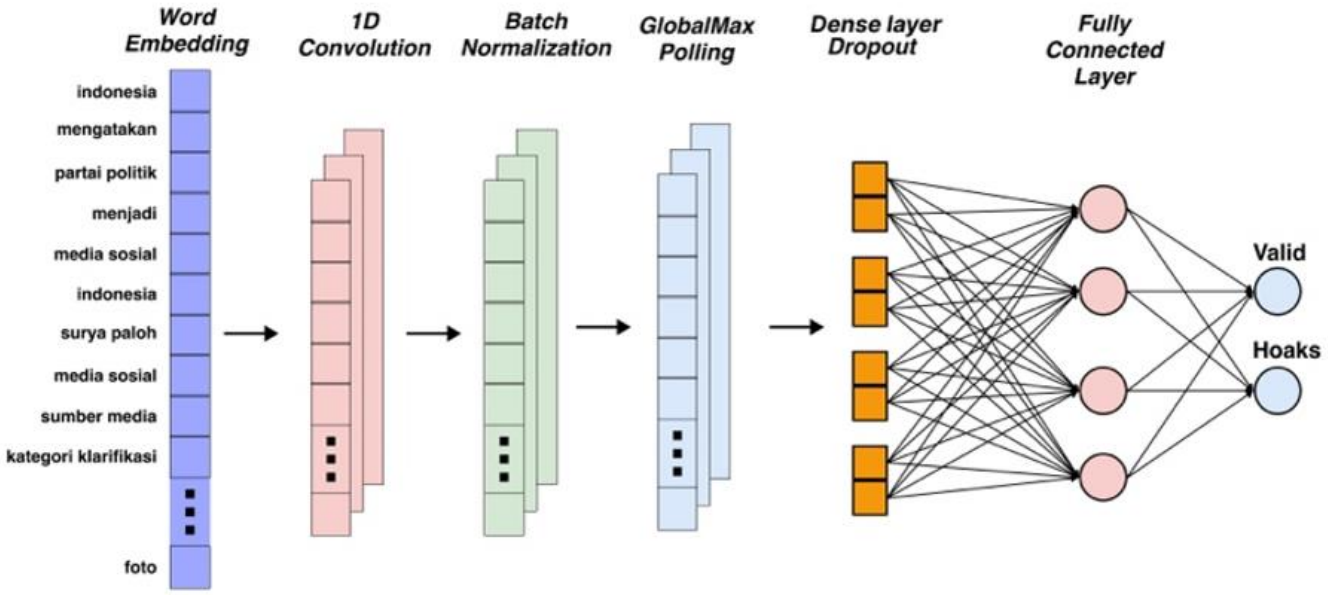
III. *ONE-DIMENSIONAL CONVOLUTIONAL NEURAL NETWORK*

1D-CNN adalah arsitektur jaringan saraf yang didesain untuk memproses data satu dimensi, seperti sinyal waktu, teks, atau data berurutan lainnya. Tidak seperti *two-dimensional* CNN (2D-CNN) yang digunakan untuk gambar, 1D-CNN lebih berfokus pada pola data linier dan sering diaplikasikan pada analisis sinyal, pemrosesan bahasa alami, dan pengenalan pola [9].

Seperti yang ditunjukkan pada Gambar 1, model 1D-CNN sangat relevan untuk mendeteksi berita hoaks karena mampu mengklasifikasikan teks ke dalam kategori tertentu, seperti hoaks atau valid [5]. Model ini dirancang untuk menangkap pola temporal atau sekuensial dalam data teks dengan mengubahnya menjadi vektor numerik, sehingga memungkinkan pemrosesan teks yang lebih efektif. Dalam pendekatan pemelajaran terbimbing, model ini dilatih menggunakan *dataset* berlabel, dengan proses pelatihan mengoptimalkan bobot jaringan berdasarkan kesalahan prediksi terhadap label yang telah diberikan, sehingga memungkinkan model mengenali karakteristik berita hoaks secara akurat [10]. Proses deteksi dilakukan dengan menganalisis struktur kalimat melalui lapisan konvolusi yang mengekstraksi fitur dari urutan kata atau token dalam teks. Lapisan ini dirancang untuk mendeteksi pola lokal, seperti susunan kata, kombinasi frasa, dan hubungan antarkata dalam konteks tertentu, yang mencakup pola linguistik dan semantik yang berguna dalam membedakan berita palsu dari berita asli. Selain itu, model ini memanfaatkan korelasi antarfitur dalam *dataset* berlabel, dengan setiap data telah dilengkapi dengan kategori berita (hoaks atau valid) [11]. Korelasi ini membantu model mengenali pola yang relevan untuk masing-masing kelas, sehingga meningkatkan akurasi deteksi serta kemampuan generalisasi terhadap data baru.

Penelitian sebelumnya telah menunjukkan bahwa CNN sangat efektif dalam analisis teks, termasuk klasifikasi berita dan tugas deteksi anomali [12]. Model 1D-CNN memiliki kemampuan dalam mengekstraksi fitur-fitur linguistik, termasuk urutan kata, pola frasa yang khas, dan karakteristik struktur sintaksis yang cenderung berulang dalam teks berita hoaks. Dengan memanfaatkan konvolusi 1D, model 1D-CNN dapat menangkap hubungan antara kata atau frasa tanpa bergantung pada urutan data secara keseluruhan.

Dibandingkan dengan *recurrent neural networks* (RNN), 1D-CNN memiliki beberapa keunggulan. RNN cenderung membutuhkan lebih banyak waktu komputasi karena mempertahankan informasi urutan kata melalui mekanisme seperti *long short-term memory* (LSTM) atau *gated recurrent unit* (GRU), yang sering meningkatkan kompleksitas model. Sebaliknya, 1D-CNN lebih efisien karena hanya mengekstrak fitur lokal melalui operasi konvolusi, sehingga meminimalkan



Gambar 1. Arsitektur model 1D-CNN.

biaya komputasi dengan tetap mempertahankan tingkat akurasi yang tinggi [3].

IV. METODOLOGI

Metode penelitian ini mencakup langkah-langkah sistematis dalam pengembangan model deteksi berita hoaks menggunakan model 1D-CNN. Tahapan penelitian diawali dengan pengumpulan data dari sumber yang relevan, diikuti oleh tahap prapemrosesan teks yang meliputi pembersihan data, *case folding*, dan penghapusan *stopword*. Data yang telah diproses kemudian dibagi menjadi data latih dan data uji untuk keperluan pelatihan serta evaluasi model.

Selanjutnya, data diubah menjadi representasi vektor numerik yang dapat diolah oleh model pemelajaran mendalam. Pada tahap pembangunan model, dilakukan pemilihan dan perancangan arsitektur 1D-CNN, termasuk penentuan lapisan konvolusi, *pooling*, dan *fully connected*. Selain itu, dilakukan penyesuaian *hyperparameter*, seperti jumlah *epoch*, ukuran *batch*, laju belajar, dan *optimizer*, untuk mengoptimalkan kinerja model.

Setelah penyesuaian *hyperparameter*, model dilatih menggunakan data latih. Proses pelatihan ini mencakup evaluasi berkala menggunakan metrik kinerja seperti akurasi, presisi, *recall*, dan *F*1-*score*. Selanjutnya, model diuji menggunakan data uji untuk menilai generalisasi dan efektivitasnya dalam mendeteksi berita hoaks. Hasil evaluasi dianalisis untuk memberikan interpretasi terhadap kinerja model serta rekomendasi perbaikan untuk penelitian selanjutnya.

A. PENGUMPULAN DATA

Pengumpulan data dalam penelitian ini didasarkan pada *dataset* berita nasional *open source* yang telah diverifikasi oleh para ahli di bidangnya. Data yang digunakan berupa artikel berita berbahasa Indonesia yang terdiri atas gabungan judul dan isi berita, dengan kategori berita valid yang diberi label 0 dan berita hoaks yang diberi label 1. Artikel berita valid diperoleh dari sumber berita nasional *open source* yang telah melalui proses pelabelan, sedangkan artikel berita hoaks diambil dari situs Turnbackhoax [13], sebuah komunitas antihoaks di Indonesia yang secara spesifik memberikan label untuk membedakan antara berita hoaks dan berita valid.

Data ini akan menjadi masukan utama untuk pelatihan dan pengujian model. Karena model 1D-CNN memproses data dalam bentuk urutan kata, teks berita diolah terlebih dahulu menggunakan teknik tokenisasi untuk mengubah setiap kata menjadi representasi numerik berupa indeks. Representasi ini memungkinkan model untuk mengenali pola-pola urutan kata yang relevan dalam teks berita, sehingga mampu melakukan klasifikasi secara efektif.

B. PRAPEMROSESAN TEKS

Prapemrosesan teks merupakan langkah penting dalam memastikan data yang digunakan bersih, relevan, dan representatif untuk menghasilkan model yang berkinerja optimal [14]. Tahap pertama dalam prapemrosesan teks adalah pembersihan data, yang mencakup penanganan nilai yang hilang (*missing values*) dengan menghapus baris yang mengandung nilai yang hilang. Selain itu, proses pembersihan data juga mencakup penghapusan data duplikat untuk mencegah bias dalam pelatihan model, lalu diakhiri dengan penggabungan data yang telah dibersihkan.

Tahap berikutnya adalah *case folding*, yaitu mengubah semua teks menjadi huruf kecil (*lowercase*) untuk memastikan konsistensi, sehingga kata-kata yang sama dengan kapitalisasi berbeda dapat dikenali sebagai entitas yang sama. Kemudian, dilakukan penghapusan simbol dan karakter khusus, seperti tanda baca, angka, atau emoji yang tidak relevan dengan konteks analisis, sehingga data menjadi lebih bersih dan terstruktur [14]. Selain itu, *stopword* seperti “dan”, “atau”, dan “yang”, yang tidak memiliki kontribusi signifikan dalam analisis semantik, dihapus untuk mengurangi dimensi data dan meningkatkan fokus pada kata-kata yang penting [14]. Dengan melakukan prapemrosesan ini, data menjadi lebih siap untuk digunakan dalam proses analisis dan pelatihan model, sehingga hasil akhirnya diharapkan lebih akurat.

C. PEMBANGUNAN DAN PELATIHAN MODEL

Pada tahap ini, proses pembangunan model 1D-CNN dimulai. Langkah pertama melibatkan tokenisasi data. Dalam hal ini, teks yang telah dipersiapkan untuk pelatihan diubah menjadi token atau suku kata yang diurutkan berdasarkan indeks kata. Proses ini menghasilkan representasi data dalam

bentuk numerik, yang kemudian dikonversi menjadi urutan agar dapat diproses oleh model. Selanjutnya, data yang telah berbentuk runtun (*sequence*) diubah menjadi format kategori untuk keperluan klasifikasi dan kemudian dibagi menjadi *dataset* pelatihan dan validasi. Langkah berikutnya adalah mendefinisikan *hyperparameter* serta merancang arsitektur model untuk memastikan bahwa model dapat mengolah data secara efisien dan mendeteksi pola-pola relevan yang mendukung tujuan klasifikasi.

Hyperparameter merupakan nilai yang ditentukan sebelum pelatihan untuk mengontrol arsitektur dan proses pemelajaran, seperti *embedding* dan filter. Model 1D-CNN dibangun dengan lapisan *embedding* untuk mengubah kata menjadi vektor, diikuti oleh *conv1d* untuk mendeteksi pola dalam urutan. Selanjutnya, *globalmaxpooling1d* digunakan untuk mereduksi dimensi, diikuti oleh lapisan *dense* [1] dan lapisan *dropout* untuk mencegah *overfitting* [15] serta diikuti dengan jumlah *epoch* = 10 [1]. Model dilatih menggunakan *hyperparameter* tersebut untuk mengoptimalkan kinerja.

D. EVALUASI MODEL

Setelah model selesai dibangun dan dilatih, evaluasi model dilakukan menggunakan metode 10-*fold cross-validation* untuk mengevaluasi kinerja model secara menyeluruh. Teknik ini dipilih karena mampu memberikan estimasi kinerja yang lebih stabil dibandingkan dengan metode pengujian tunggal serta mengurangi risiko *overfitting* dengan menguji model di berbagai subset data [16]. Pada setiap iterasi, *dataset* dibagi menjadi sepuluh subset, dengan sembilan subset digunakan untuk pelatihan dan satu subset untuk pengujian. Proses ini diulang hingga semua subset telah digunakan sebagai data uji. Hasil prediksi model dibandingkan dengan label asli untuk menghitung metrik evaluasi, seperti akurasi, presisi, *recall*, dan *F*1-*score* [17]. Metrik ini memberikan gambaran lengkap tentang kemampuan model, mulai dari tingkat ketepatan dalam memprediksi hingga sensitivitasnya terhadap data hoaks. Evaluasi ini tidak hanya penting untuk menilai kemampuan generalisasi model terhadap data baru, tetapi juga untuk mengidentifikasi potensi perbaikan pada struktur atau proses pelatihan model. Jika ditemukan adanya ketidakseimbangan kinerja pada kategori tertentu, penyesuaian dapat dilakukan, seperti menambahkan data pada kategori yang kurang terwakili atau mengoptimalkan parameter *hyperparameter* model. Langkah ini memastikan bahwa model yang dihasilkan memiliki kinerja yang andal dan siap diterapkan pada skenario dunia nyata.

V. HASIL DAN PEMBAHASAN

Bagian pembahasan ini dibagi menjadi empat bagian utama. Bagian pertama menjelaskan proses hasil pengumpulan *dataset*, termasuk sumber data dan kriteria pemilihan berita. Bagian kedua membahas hasil prapemrosesan teks, seperti pembersihan data, *case folding*, dan penghapusan *stopword*. Bagian ketiga menjelaskan pelatihan dan arsitektur model, termasuk pemilihan *hyperparameter* dan hasil pelatihan dan bagian keempat membahas hasil evaluasi model menggunakan metrik seperti akurasi, presisi, *recall*, dan *F*1-*score* untuk mengukur efektivitas deteksi berita hoaks.

A. HASIL PENGUMPULAN DATA

Pengumpulan data dalam penelitian ini mengacu pada *dataset* berita nasional *open source* serta *dataset* gratis yang telah diverifikasi oleh para ahli di bidangnya. Keseluruhan data yang digunakan berjumlah 12.151 data berita, yang diperoleh melalui dua tahap utama. Proses ini dirancang untuk memastikan bahwa *dataset* yang digunakan bersifat representatif, berkualitas, dan mendukung tujuan penelitian, yaitu membangun model deteksi berita hoaks dengan kinerja dan akurasi tinggi. Selain itu, penggunaan *dataset* yang besar diharapkan dapat meningkatkan kemampuan generalisasi model terhadap berbagai jenis berita.

Pada tahap awal, diperoleh sebanyak 1.246 data berita berbahasa Indonesia yang diambil dari penelitian sebelumnya [18]. *Dataset* ini terdiri atas 553 berita valid dan 693 berita hoaks. Proses seleksi dan verifikasi dilakukan untuk memastikan kualitas data, termasuk validitas, relevansi, dan keseimbangan jumlah antara berita valid dan berita hoaks. *Dataset* awal ini digunakan sebagai dasar untuk eksperimen awal dalam pengembangan model deteksi berita hoaks.

Pada tahap kedua, *dataset* diperluas secara signifikan dengan menambahkan 10.905 data tambahan yang berasal dari *dataset* penelitian terdahulu dan diperoleh melalui proses *scraping* dari berbagai portal berita, sebagaimana dijelaskan dalam penelitian sebelumnya [13]. Proses *scraping* dilakukan secara sistematis dengan merujuk pada dua kategori sumber utama, yaitu situs berita terpercaya dan portal berita hoaks. Penambahan *dataset* ini bertujuan untuk meningkatkan keberagaman dan cakupan data, sehingga dapat memperkuat kemampuan model dalam mendeteksi berita palsu secara lebih akurat.

Berita valid dikumpulkan dari portal berita terkemuka di Indonesia, seperti Kompas dan CNN, yang memiliki reputasi tinggi dalam menyajikan informasi faktual. Sementara itu, berita hoaks diperoleh dari situs Turnbackhoax, yang secara khusus mendokumentasikan dan memverifikasi berbagai informasi palsu yang beredar di masyarakat. Diperoleh hasil bahwa *dataset* yang terkumpul terdiri atas 5.276 berita valid yang diberi label 0 dan 6.875 berita hoaks yang diberi label 1, sehingga memastikan keseimbangan data yang baik untuk pelatihan model deteksi berita hoaks.

B. PRAPEMROSESAN TEKS

Tahapan prapemrosesan teks diawali dengan proses pembersihan data, yaitu membersihkan data dari nilai yang tidak lengkap atau *missing values*. *Dataset* yang bersumber dari Kompas memiliki 27 nilai yang tidak diketahui, sehingga dilakukan penghapusan baris yang mengandung *missing values* pada kolom terkait. Setelah proses ini, jumlah keseluruhan data yang tersisa dari sumber Kompas menjadi 4.723 baris.

Dataset dari Turnbackhoax juga mengalami permasalahan serupa, dengan sejumlah baris data memiliki nilai tidak lengkap pada kolom tertentu. Oleh karena itu, dilakukan penghapusan baris yang tidak memiliki data lengkap untuk meningkatkan kualitas *dataset*. Setelah proses pembersihan data, jumlah keseluruhan data yang tersedia dari Turnbackhoax mencapai 6.592 baris, yang terdiri atas beberapa label, yaitu Misinformasi, Hoaks, Benar, dan Salah. Selanjutnya, dilakukan proses penyaringan label pada *dataset* dari Turnbackhoax untuk memastikan keseragaman kategori data. Baris dengan label Disinformasi, Salah, dan Misinformasi dipertahankan, sedangkan baris yang berlabel Benar dihapus agar *dataset* hanya berisi berita dengan label Hoaks. Setelah tahap ini, jumlah keseluruhan data yang telah dibersihkan dan siap digunakan menjadi 6.182 baris.

Pada *dataset* yang berasal dari penelitian sebelumnya [18], dilakukan perubahan penyamaan label, dengan data yang berlabel hoaks diberi label 1 dan data yang berlabel valid diberi

label 0. Jumlah data berita valid sebanyak 553 dan berita hoaks sebanyak 693. Setelah seluruh data dinyatakan bersih, dilakukan proses penggabungan terhadap data yang telah melalui tahap pembersihan. Namun, data tersebut masih belum sepenuhnya siap untuk digunakan dalam pelatihan model, sehingga perlu melalui tahap prapemrosesan lanjutan sebelum digunakan sebagai data latih. Pada *dataset* tersebut, terdapat 12.151 data, yang terdiri atas 6.874 data berita hoaks (56,5%) dan 5.276 data berita valid (43,5%). Data ini mencerminkan proporsi yang tidak terlalu jauh, seperti terlihat pada Tabel I.

Pada Tabel I, data yang telah dibersihkan dan digabungkan masih memiliki teks yang tidak konsisten, seperti terdapat penggunaan huruf kapital, tanda baca, dan terdapat beberapa simbol yang tidak diperlukan. Langkah selanjutnya adalah *case folding*, yaitu mengonversi seluruh teks menjadi huruf kecil untuk memastikan konsistensi, sehingga kata seperti "Indonesia" dan "indonesia" tidak dianggap berbeda oleh model. Selain itu, teks dibersihkan dari simbol nonalfabet yang dapat mengganggu pemahaman pola dalam data. Setelah *case folding*, dilakukan penghapusan *stopword*, yaitu menghapus kata-kata umum seperti "dan", "atau", dan "yang", yang tidak memiliki kontribusi signifikan dalam analisis. Proses ini menggunakan pustaka Natural Language Toolkit (NLTK) dan Sastrawi, sehingga teks yang tersisa hanya mengandung kata-kata esensial yang membantu model mengenali konteks secara lebih efektif [19].

Dataset hasil prapemrosesan disajikan pada Tabel II, yang menunjukkan perubahan signifikan dalam struktur data, memastikan bahwa teks yang digunakan bebas dari elemen-elemen tidak relevan seperti simbol, angka, atau kata-kata umum yang tidak memiliki kontribusi signifikan dalam analisis. Untuk memberikan visualisasi tambahan, dibuat juga *word cloud* yang ditampilkan pada Gambar 2. Langkah-langkah ini memastikan bahwa data yang digunakan memenuhi standar kualitas yang diperlukan untuk analisis lanjutan dan pengembangan model pemelajaran mendalam.

Pada Tabel II, *stopword*, yaitu "dan", "atau", "dengan", dan kata-kata umum lainnya, dihapus dari atribut teks menggunakan pustaka Sastrawi. Penghapusan *stopword* bertujuan untuk menghilangkan kata-kata yang tidak memiliki makna signifikan dalam analisis, sehingga hanya menyisakan kata-kata kunci yang relevan. Proses ini dilakukan untuk meningkatkan kualitas data pada kolom "Teks" guna mendukung analisis lanjutan, seperti deteksi hoaks. Gambar 2 merupakan *word cloud* yang memvisualisasikan kata-kata yang paling sering muncul dalam *dataset* setelah prapemrosesan, memberikan wawasan awal mengenai pola umum, tema, dan istilah yang dominan dalam data [20].

C. TRAINING MODEL

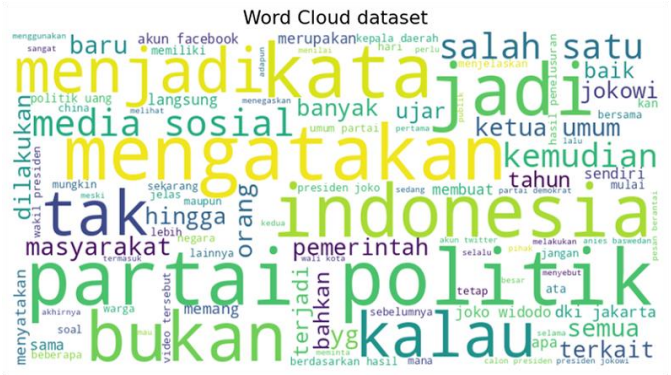
Setelah teks melalui tahap prapemrosesan, data siap untuk digunakan dalam pelatihan model klasifikasi, khususnya untuk deteksi hoaks. Proses ini dimulai dengan mengubah teks menjadi format kategorial yang sesuai untuk model. Data teks diambil dari *dataset* untuk dianalisis lebih lanjut melalui tahap tokenisasi, *padding*, dan transformasi lainnya agar dapat diproses oleh model pemelajaran mendalam. Label yang menunjukkan kategori berita (seperti "hoaks" dan "valid") kemudian diubah ke dalam format biner menggunakan *one-hot encoding*. Format ini mengonversi label teks menjadi bentuk seperti [1, 0] untuk berita valid dan [0, 1] untuk berita hoaks. Tabel III menyajikan contoh hasil proses ini berikut beberapa contoh teks berita beserta label kategorinya.

TABEL I
DATASET YANG TELAH MELALUI FASE PEMBERSIHAN DATA

Label	Teks
1	Anies Baswedan Dekat Dengan Aliran Krsiten Se...
1	Hakim Wahyu Iman Santoso Alami Kecelakaan Tun...
1	GAMBAR MEGAWATI DAN PUAN BERMAIN SLOT Nenek l...
…	…
0	Efek Ekor Jas Pencalonan Anies, Elektabilitas ...
0	Survei Litbang Kompas: PDI-P, Gerindra, dan Go...

TABEL II
DATASET YANG TELAH MELALUI FASE PENGHAPUSAN STOPWORD

Label	Teks
1	anies baswedan dekat aliran krsiten sesat alph...
1	hakim wahyu iman santoso alami kecelakaan tung...
1	gambar megawati puan bermain slot nenek lampir...
1	jonathan latumahina seorang nasrani penyusup o...
0	kemenangan telak anies sandiaga pilkada dki...
...	...
0	pesawat hercules milik tni au pesawat hercules chs milik tni angkatan udara...



Gambar 2. *Word cloud* dari *dataset*.

Proses tokenisasi merupakan langkah penting dalam pengolahan data teks. Teks diubah menjadi urutan token yang merepresentasikan kata atau subkata dalam kosakata (*vocabulary*) yang telah ditentukan sebelumnya. Selama proses ini, kata-kata yang tidak ditemukan dalam kosakata (*out-of-vocabulary*/OOV) akan digantikan oleh token khusus yang dirancang untuk menangani kasus tersebut. Tokenisasi dilakukan menggunakan *tokenizer* bawaan dari pustaka Tensorflow Keras, yang menyediakan fleksibilitas dalam pengaturan parameter seperti *num_words* (jumlah kata maksimum yang dipertimbangkan dalam kosakata) dan *oov_token* untuk mengidentifikasi kata-kata yang tidak dikenal. Proses ini bertujuan untuk memastikan bahwa semua teks dapat diubah menjadi representasi numerik yang dapat diproses oleh model secara konsisten, termasuk untuk data dengan kata-kata yang jarang muncul.

Setelah *tokenisasi*, diterapkan proses *padding* menggunakan fungsi *pad_sequences*. *Padding* bertujuan untuk menyamakan panjang setiap urutan teks agar model dapat memproses masukan dengan ukuran yang konsisten. Teks yang lebih pendek dari panjang maksimum akan dilengkapi dengan nilai nol, sementara teks yang lebih panjang akan dipotong sesuai dengan parameter *trunc_type* yang telah ditentukan, seperti "*pre*" atau "*post*". Proses ini sangat penting untuk memastikan model menerima masukan dalam bentuk yang seragam, mempermudah pelatihan, dan meningkatkan efisiensi.

TABEL III
HASIL *DATASET* SETELAH DIUBAH MENJADI KATEGORIAL

Teks	Label	Keterangan
luhut curigai motif politik balik penolakan wni wuhan natuna adanya penolakan masyarakat kabupaten natuna kepulauan riau warga negara …..	[*True False*]	Teks tersebut merupakan salah satu contoh berita valid, sehingga label bernilai *True False*
sepatu matamata perang dunia ii sepatu dipakai matamata sekutu selama perang dunia ii menyesatkan jerman mengarahkan kearah berlawanan ….	[*False True*]	Teks tersebut merupakan salah satu contoh berita hoaks, sehingga label bernilai *False True*

Setelah melalui tahap tokenisasi dan *padding*, *dataset* dibagi menjadi dua bagian, yaitu data latih (80%) dan data uji (20%). Data latih digunakan untuk melatih model, sementara data uji digunakan untuk mengevaluasi kinerja model selama proses pelatihan. Evaluasi ini membantu dalam memilih model terbaik berdasarkan metrik seperti akurasi dan *loss*. Dengan membagi *dataset*, model dapat diuji pada data yang belum pernah dilihat untuk memastikan generalisasi yang baik.

Tahap akhir dalam proses ini adalah membangun arsitektur model 1D-CNN yang dirancang untuk mendeteksi berita hoaks dengan mengidentifikasi pola-pola unik dalam teks berita [9]. Model ini memanfaatkan lapisan konvolusi untuk mengekstraksi fitur penting dari data teks, yang kemudian diteruskan ke lapisan *dense* untuk pengambilan keputusan akhir. *Hyperparameter* seperti ukuran *embedding*, jumlah filter, ukuran *kernel*, fungsi aktivasi, laju belajar, *dropout rate*, dan fungsi *loss* telah disesuaikan untuk mengoptimalkan kinerja model dan mencegah *overfitting*.

Model dimulai dengan lapisan *embedding*, yang berfungsi untuk mengubah teks menjadi representasi vektor berdimensi tertentu [21]. Representasi ini memungkinkan teks diubah ke dalam bentuk yang dapat dipahami oleh model dengan tetap mempertahankan makna kontekstualnya. Setelah itu, lapisan *conv1d* digunakan untuk mengekstraksi fitur-fitur penting dari urutan token, yang memungkinkan model mengenali pola-pola dalam data teks dan memahami hubungan antarkata dalam konteks tertentu.

Guna meningkatkan stabilitas dan efisiensi pelatihan, model dilengkapi dengan lapisan *batch normalization* yang berfungsi untuk menormalkan keluaran dari lapisan sebelumnya. Proses ini membantu mengurangi pergeseran distribusi data selama pelatihan, mempercepat konvergensi, dan meningkatkan kinerja model [22]. Setelah itu, diterapkan lapisan *globalmaxpooling1d* untuk mereduksi dimensi data dengan mengambil nilai maksimum dari seluruh panjang urutan sebagai fitur yang paling dominan [23]. Lapisan ini tidak hanya membantu dalam mengurangi kompleksitas data, tetapi juga mempertahankan informasi penting dari fitur utama yang relevan dalam mendeteksi pola. Kombinasi ini memastikan bahwa model dapat menangkap representasi data yang lebih informatif dan stabil, sehingga meningkatkan akurasi dalam klasifikasi berita hoaks.

Lapisan terakhir berupa lapisan *dense* dengan dua unit keluaran, yang menggunakan fungsi aktivasi *softmax* untuk mengklasifikasikan teks ke dalam dua kategori, yaitu hoaks atau valid. Untuk meningkatkan kemampuan generalisasi model dan mencegah *overfitting*, ditambahkan lapisan *dropout* dengan porsi 0,5. Teknik ini secara acak menonaktifkan sejumlah neuron selama proses pelatihan, sehingga mencegah model terlalu bergantung pada neuron tertentu. Dengan demikian, lapisan *dropout* membantu model untuk belajar pola yang lebih umum, mengurangi risiko *overfitting*, dan meningkatkan kemampuannya dalam menggeneralisasi pada data yang belum pernah dilihat sebelumnya. Penambahan lapisan *dropout* ini diharapkan dapat meningkatkan stabilitas dan akurasi model dalam deteksi berita hoaks, sehingga menghasilkan prediksi yang lebih *robust* dan dapat diandalkan pada berbagai kondisi data [24].

Proses pelatihan model 1D-CNN untuk mendeteksi berita hoaks dilakukan dengan menggunakan *dataset* yang telah diproses sebelumnya. Pelatihan dilakukan selama sepuluh *epoch* [1]. Hasil dari proses pelatihan ini ditampilkan pada Gambar 3, yang menunjukkan metrik kinerja seperti akurasi dan nilai *loss* pada data pelatihan dan validasi di setiap *epoch*. Analisis terhadap grafik tersebut memberikan gambaran mengenai konvergensi model dan kinerjanya dalam mendeteksi berita hoaks berdasarkan *dataset* yang digunakan.

Tampak pada Gambar 3 bahwa pada *epoch* pertama, model mencapai akurasi pelatihan sebesar 88,41% dan akurasi validasi 97,20%, yang menunjukkan kemampuan awal yang baik dalam mempelajari pola *dataset*. Akurasi pelatihan terus meningkat hingga 98,16%, sedangkan akurasi validasi stabil di 97,98%. Penurunan *training loss* dari 0,2533 menjadi 0,0408 menandakan proses pemelajaran yang efektif, meskipun fluktuasi *validation loss* dari 0,3171 ke 0,1134 menunjukkan potensi *overfitting* ringan. Dengan durasi rata-rata 9 detik per *epoch*, model 1D-CNN ini terbukti efisien dan memiliki kemampuan generalisasi yang baik.

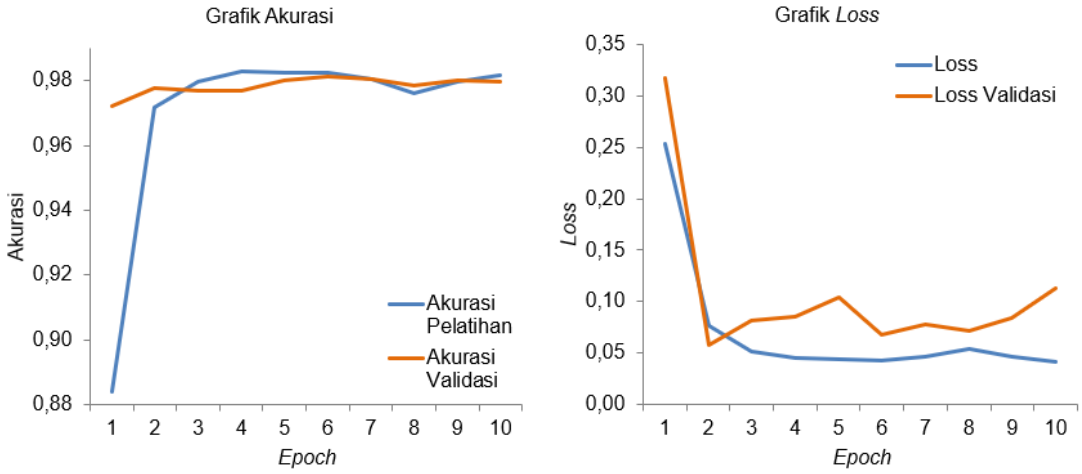
Hasil pelatihan menunjukkan kinerja yang lebih baik dan seimbang. Arsitektur 1D-CNN dengan *batch normalization* tanpa menggunakan lapisan *dropout* [1] mencapai 100% akurasi pada data pelatihan. Replikasi hasil penelitian tersebut ditampilkan pada Gambar 4.

Gambar 4 menunjukkan bahwa model mencapai akurasi pelatihan sebesar 100% dan akurasi validasi sebesar 98%. Perbedaan antara akurasi pelatihan dan akurasi validasi serta selisih antara *training loss* dan *validation loss* relatif kecil. Namun, akurasi pelatihan yang sempurna dapat mengindikasikan potensi *overfitting*, yang berarti kinerja model pada data uji mungkin tidak optimal [24].

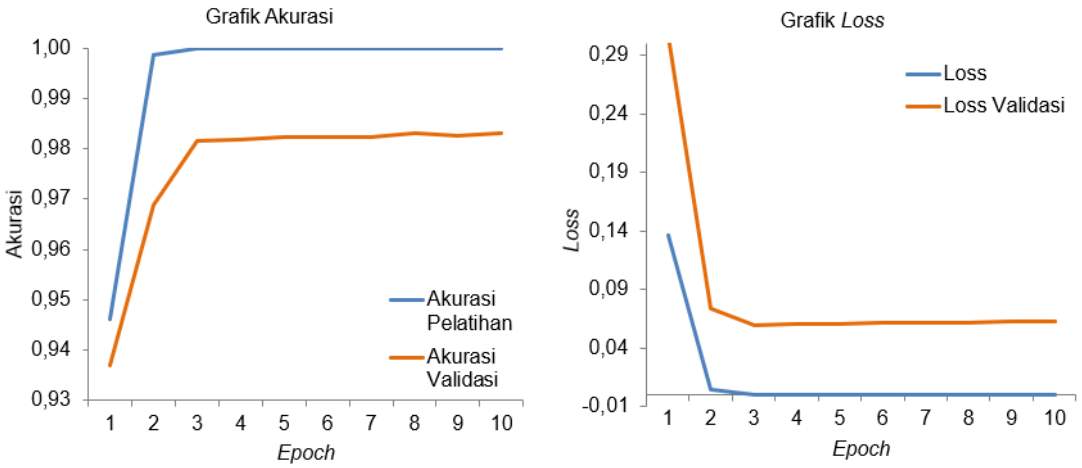
Hasil ini berbeda dengan pelatihan sebelumnya yang menerapkan *dropout*, dengan akurasi pelatihan sebesar 98,16% dan akurasi validasi tetap stabil di 97,98%. Perbedaan ini menunjukkan bahwa penggunaan *dropout* dapat membantu mengurangi risiko *overfitting*. Oleh karena itu, diperlukan evaluasi lebih lanjut atau penerapan teknik regularisasi, seperti *dropout* atau pengurangan kompleksitas model, untuk meningkatkan kemampuan generalisasi model.

D. EVALUASI MODEL

Model dievaluasi menggunakan metode 10-*fold cross-validation* untuk mengukur konsistensi kinerja pada berbagai subset data [1]. Dalam metode ini, *dataset* dibagi menjadi sepuluh *fold* yang seimbang, dengan setiap *fold* secara bergantian menjadi data validasi, sedangkan semiblan *fold* lainnya digunakan untuk pelatihan. Pendekatan ini efektif dalam mengurangi bias, meningkatkan reliabilitas hasil evaluasi, dan memberikan gambaran yang lebih akurat mengenai kemampuan model dalam menghadapi data baru.



Gambar 3. Grafik hasil pelatihan model.



Gambar 4. Grafik hasil pelatihan menggunakan arsitektur model penelitian [1].

Hasil evaluasi kinerja model dianalisis menggunakan *confusion matrix* untuk menghitung metrik utama, termasuk akurasi, presisi, *recall*, *dan F*1*-score*. Metrik akurasi mengukur proporsi total prediksi yang benar, sedangkan presisi menghitung proporsi prediksi positif yang benar-benar relevan. *Recall* digunakan untuk mengevaluasi kemampuan model dalam mengidentifikasi seluruh sampel positif yang sebenarnya dan *F*1*-score* merupakan metrik harmonis yang menggabungkan presisi dan *recall* secara seimbang. Analisis ini memberikan wawasan mendalam tentang kekuatan dan kelemahan model dalam mendeteksi berita hoaks dengan tingkat validitas yang tinggi. Berdasarkan *confusion matrix*, metrik-metrik ini dapat dihitung menggunakan (1) hingga (4).

$$Akurasi = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

$$Presisi = \frac{TP}{TP+FP} \quad (2)$$

$$Recall = \frac{TP}{TP+FN} \quad (3)$$

$$F1-Score = \frac{2 \cdot Presisi \cdot Recall}{Presisi+Recall}. \quad (4)$$

Selain menggunakan *confusion matrix*, hasil evaluasi juga dibandingkan dengan penelitian sebelumnya untuk menilai keunggulan atau peningkatan kinerja model. Analisis ini memberikan wawasan yang mendalam mengenai kekuatan dan kelemahan model dalam mendeteksi berita hoaks secara efektif serta menunjukkan relevansi metode yang digunakan dengan pendekatan yang telah ada. Hasil evaluasi model disajikan pada Tabel IV.

Pendekatan ini penting dilakukan untuk memastikan bahwa model yang dikembangkan tidak hanya memiliki kinerja yang baik secara teoretis, tetapi juga menunjukkan keunggulan yang nyata jika dibandingkan dengan metode lain yang telah digunakan sebelumnya. Sebagai contoh, penelitian terdahulu yang menggunakan pendekatan *naïve Bayes* atau SVM mungkin memiliki tingkat akurasi yang memadai, tetapi cenderung menghadapi kendala dalam mengatasi kompleksitas data teks yang sangat besar. Dengan menggunakan model 1D-CNN dalam penelitian ini, terdapat potensi untuk meningkatkan kemampuan model dalam memahami pola-pola teks secara lebih mendalam, khususnya dalam mendeteksi berita palsu atau hoaks yang sering kali memiliki struktur kalimat yang tidak teratur.

Analisis perbandingan dengan penelitian sebelumnya difokuskan pada metrik evaluasi, seperti akurasi, presisi, dan *recall* guna memberikan penilaian yang objektif terhadap kinerja model dalam klasifikasi berita hoaks. Pendekatan ini memungkinkan evaluasi yang terukur dan relevan secara langsung dengan tujuan utama penelitian, yakni meningkatkan kinerja model deteksi berita hoaks berbasis teks. Hal ini disebabkan struktur 1D-CNN dirancang untuk memproses data sekuensial dengan lebih efisien melalui penggunaan filter konvolusi yang secara langsung dapat menangkap fitur lokal dari teks masukan.

Selain itu, penelitian ini juga mengevaluasi kemampuan model dalam menangani berbagai jenis berita palsu. Salah satu tantangan utama dalam mendeteksi berita hoaks adalah adanya variasi gaya penulisan yang digunakan oleh pembuat hoaks untuk menyamarkan informasi palsu. Dalam hal ini, pendekatan 1D-CNN memiliki keunggulan dalam mengenali pola-pola tertentu pada level karakter maupun kata. Hal ini memungkinkan model untuk mendeteksi hoaks yang memiliki pola penulisan serupa dengan data pelatihan, meskipun menggunakan kosakata yang berbeda.

Dalam penelitian sebelumnya, salah satu kendala yang sering dihadapi adalah rendahnya kinerja model ketika diterapkan pada data yang memiliki distribusi yang berbeda dari data pelatihan. Untuk mengatasi hal ini, model dalam penelitian ini dirancang dengan menggunakan teknik regularisasi seperti *dropout* dan *batch normalization* untuk mengurangi risiko *overfitting*.

Hasil evaluasi yang disajikan pada Tabel IV menunjukkan bahwa model 1D-CNN ini tidak hanya memiliki akurasi yang lebih tinggi, tetapi juga kinerja yang konsisten pada metrik lain, seperti presisi dan *recall*, dibandingkan dengan penelitian sebelumnya. Dalam konteks deteksi berita palsu, presisi yang tinggi menunjukkan bahwa model memiliki kemampuan yang baik dalam meminimalkan *false positive*, yaitu berita yang sebenarnya valid tetapi terdeteksi sebagai hoaks. Sebaliknya, *recall* yang tinggi menunjukkan bahwa model mampu mendeteksi sebagian besar berita hoaks dalam *dataset*, sehingga risiko melewatkan informasi yang salah dapat diminimalkan. Dengan membandingkan hasil penelitian ini dengan pendekatan lain yang telah diterapkan dalam berbagai konteks, dapat disimpulkan bahwa model 1D-CNN yang diusulkan memiliki potensi besar untuk menjadi solusi yang efektif dalam mendeteksi berita hoaks secara otomatis. Analisis ini tidak hanya memberikan kontribusi terhadap literatur akademik, tetapi juga membuka peluang untuk pengembangan lebih lanjut, termasuk penerapan pada platform yang lebih luas seperti media sosial.

Hasil evaluasi rata-rata menunjukkan akurasi 97,74% dengan simpangan baku ±0,0040. Nilai presisi, *recall*, dan *F*1-*score* juga konsisten tinggi, seperti ditunjukkan pada Tabel IV. Hal ini membuktikan kemampuan generalisasi model yang baik. Selain itu, kinerja model dalam penelitian ini melampaui hasil penelitian sebelumnya [6] yang menggunakan arsitektur 1D-CNN dengan teknik *dropout* serta model gabungan 1D-CNN dan BiLSTM untuk deteksi berita hoaks. Pada penelitian tersebut, model 1D-CNN mencapai akurasi sebesar 96%. Dengan demikian, pendekatan yang diusulkan dalam penelitian ini tidak hanya berhasil mengatasi risiko *overfitting*, tetapi juga menunjukkan keunggulan dalam mendeteksi berita hoaks dengan akurasi yang lebih tinggi. Perbandingan kinerja model deteksi berita hoaks dengan penelitian sebelumnya ditunjukkan pada Tabel V.

Berdasarkan Tabel V, dapat dilihat bahwa model 1D-CNN + *batch normalization* + *dropout* yang diusulkan dalam penelitian ini unggul dibandingkan dengan model lainnya dalam konteks deteksi berita hoaks. Model ini tidak hanya menunjukkan kinerja tertinggi yang didasarkan pada akurasi tertinggi.

Keunggulan model ini dapat dijelaskan oleh kemampuan 1D-CNN untuk menangkap pola linguistik mendalam dalam data teks, yang dikombinasikan dengan penggunaan *batch normalization* dan teknik *dropout* yang efektif dalam mengurangi *overfitting* [24]. Selain itu, pendekatan ini memanfaatkan efisiensi komputasi 1D-CNN, sehingga lebih cepat dalam pelatihan dibandingkan model gabungan CNN-BiLSTM. Secara keseluruhan, penelitian ini membuktikan bahwa pendekatan 1D-CNN + *batch normalization* + *dropout* tidak hanya unggul dalam hal akurasi, tetapi juga memberikan solusi yang lebih sederhana, efisien, dan efektif dalam mendeteksi berita hoaks. Hasil ini memberikan kontribusi penting dalam pengembangan sistem deteksi berita hoaks berbasis NLP yang dapat diimplementasikan secara praktis di era digital, sekaligus meningkatkan literasi informasi masyarakat. Dengan mengadaptasi model ini pada konteks bahasa Indonesia, penelitian ini juga membuka potensi penerapan serupa dalam berbagai bahasa dan bidang lainnya serta memberikan dasar yang kuat untuk riset lebih lanjut dalam mendeteksi hoaks secara otomatis dan tepat waktu.

TABEL IV
HASIL EVALUASI MODEL

Metrik	*Fold 1*	*Fold 2*	…	*Fold 10*	Rata-Rata
Akurasi	0,9737	0,9786	…	0,9712	0,9774 (± 0,0040)
Presisi	0,9740	0,9788	…	0,9712	0,9775 (± 0,0039)
Recall	0,9737	0,9786	…	0,9712	0,9774 (± 0,0040)
*F*1-*Score*	0,9736	0,9786	…	0,9712	0,9773 (± 0,0040)

TABEL V
TABEL KOMPARASI DENGAN PENELITIAN SEBELUMNYA

Model	Akurasi (%)	Presisi (%)	*Recall* (%)
1D-CNN+ *Batch Normalization* +*Dropout*	97,74	97,75	97,74
BI-LSTM + *Dropout* [1]	96,60	96,56	96,54
Gabungan CNN-SVM [11]	95,99	-	-
Naïve Bayes [1]	86,20	86,95	86,20
Gabungan CNN-BI LSTM [6]	95,94	-	-
GRU [14]	90,00	90,00	87,00

VI. KESIMPULAN

Model 1D-CNN menunjukkan kinerja yang sangat baik dalam mendeteksi dan mengklasifikasikan berita hoaks, dengan rata-rata akurasi sebesar 0,9774, presisi sebesar 0,9775, *recall* sebesar 0,9774, dan *F*1-*score* sebesar 0,9773. Evaluasi ini dilakukan menggunakan 10-*fold cross-validation* pada *dataset* yang terdiri atas 12.151 artikel berita, dengan distribusi 56% berita hoaks dan 44% berita valid. Kinerja ini mencerminkan kemampuan model, tidak hanya dalam mendeteksi keberadaan berita hoaks, tetapi juga mengklasifikasikan berita menjadi kategori hoaks dan valid secara konsisten dan akurat. Proses deteksi dilakukan dengan mengidentifikasi pola linguistik dan semantik melalui lapisan 1D-CNN, yang mampu mengekstraksi fitur seperti susunan kata dan kombinasi frasa untuk menghasilkan prediksi yang tepat. Korelasi antara kemampuan deteksi dan klasifikasi sangat erat, dengan akurasi model dalam mendeteksi berita hoaks bergantung pada efisiensinya dalam mengenali pola unik dari setiap kategori berita. Penelitian selanjutnya disarankan untuk mengintegrasikan teknik *stemming* dan *lemmatization* pada tahap prapemrosesan guna meningkatkan representasi teks, serta memperhatikan distribusi *dataset* berdasarkan kategori berita, seperti politik, teknologi, hiburan, pendidikan, dan olahraga, untuk memastikan model dapat menangkap pola kontekstual yang lebih luas dan seimbang antar kategori-

kategori berita. Optimalisasi model lebih lanjut diharapkan mampu meningkatkan kinerja model. Implementasi model dalam bentuk aplikasi Android atau situs web untuk deteksi hoaks tidak hanya bertujuan untuk mempertahankan kinerja model yang tinggi, tetapi juga memastikan bahwa model tersebut dapat diintegrasikan dan digunakan secara nyata oleh masyarakat.

KONTRIBUSI PENULIS

Konseptualisasi, Muhammad Zuama Al Amin dan Muhammad ‘Ariful Furqon; metodologi, Muhammad Zuama Al Amin; perangkat lunak, Muhammad Zuama Al Amin; validasi, Muhammad Zuama Al Amin, Muhammad ‘Ariful Furqon, dan Dwi Wijonarko; analisis formal, Muhammad Zuama Al Amin; investigasi, Muhammad Zuama Al Amin; sumber daya, Muhammad Zuama Al Amin; kurasi data, Muhammad Zuama Al Amin; penulisan—penyusunan draf asli, Muhammad Zuama Al Amin; penulisan—peninjauan dan penyuntingan, Muhammad Zuama Al Amin; visualisasi, Muhammad Zuama Al Amin; pengawasan, Muhammad Zuama Al Amin; administrasi proyek, Muhammad Zuama Al Amin; akuisisi pendanaan, Muhammad ‘Ariful Furqon.

UCAPAN TERIMA KASIH

Ucapan terima kasih disampaikan kepada semua pihak yang telah memberikan dukungan dalam penyusunan penelitian ini. Secara khusus, penulis menyampaikan apresiasi kepada Fakultas Ilmu Komputer, Universitas Jember, atas fasilitas dan dukungan yang diberikan selama penelitian ini berlangsung. Penulis juga berterima kasih kepada para kolega dan pembimbing atas masukan berharga yang telah membantu dalam penyusunan naskah ini. Selain itu, apresiasi yang mendalam disampaikan kepada keluarga dan teman-teman atas doa dan motivasi yang tiada henti. Terakhir, penulis mengucapkan terima kasih kepada semua pihak yang secara langsung maupun tidak langsung telah berkontribusi dalam penelitian ini.

REFERENSI

[1] B.P. Nayoga, R. Adipradana, R. Suryadi, dan D. Suhartono, “Hoax analyzer for Indonesian news using deep learning models,” *Procedia Comput. Sci.*, vol. 179, hal. 704–712, Feb. 2021, doi: 10.1016/j.procs.2021.01.059.

[2] Mukhlizar dan F. Yuliani, “Hoax dan tanggung jawab sosial pengguna sosial media,” *Kaganga, J. Pendidik. Sej. Ris. Sos. Hum.*, vol. 6, no. 2, hal. 461–466, Jul.-Des. 2023, doi: 10.31539/kaganga.v6i2.7695.

[3] T.C. Praha, Widodo, dan Murien Nugraheni, “Indonesian fake news classification using transfer learning in CNN and LSTM,” *JOIV, Int. J. Inform. Visualization*, vol. 8, no. 3, hal. 1213–1220, Sep. 2024, doi: 10.62527/joiv.8.2.2126.

[4] D. Amelia dkk., “Variabel yang memengaruhi kemampuan literasi membaca siswa Indonesia: Analisis berdasarkan pendekatan MARS,” *J. Pendidik. Kebud.*, vol. 9, no. 2, hal. 205–217, Des. 2024, doi: 10.24832/jpnk.v9i2.4966.

[5] J.C.S. Reis dkk., “Supervised learning for fake news detection,” *IEEE Intell. Syst.*, vol. 34, no. 2, hal. 76–81, Mar./Apr. 2019, doi: 10.1109/MIS.2019.2899143.

[6] A.D. Safira dan E.B. Setiawan, “Hoax detection in social media using bidirectional long short-term memory (Bi-LSTM) and 1 dimensional-convolutional neural network (1D-CNN) methods,” dalam *2023 11th Int. Conf. Inf. Commun. Technol. (ICoICT)*, 2023, hal. 355–360, doi: 10.1109/ICoICT58202.2023.10262528.

[7] M.F. Mridha dkk., “A comprehensive review on fake news detection with deep learning,” *IEEE Access*, vol. 9, hal. 156151–156170, Nov. 2021, doi: 10.1109/ACCESS.2021.3129329.

[8] M.G. Ayu, “Peneliti manfaatkan AI untuk deteksi berita palsu COVID-19.” Cloud Computing Indonesia. Access date: 20-Nov-2024. [Online]. Tersedia: https://www.cloudcomputing.id/berita/peneliti-manfaatkan-ai-dalam-mendeteksi-berita-palsu

[9] S. Kiranyaz dkk., “1D convolutional neural networks and applications: A survey,” *Mech. Syst. Signal Process.*, vol. 151, hal. 1–21, Apr. 2021, doi: 10.1016/j.ymssp.2020.107398.

[10] R. Oshikawa, J. Qian, dan W. Y. Wang, “A survey on natural language processing for fake news detection,” dalam *Proc. 12th Lang. Resour. Eval. Conf.*, 2020, hal. 6086–6093.

[11] M. Benedict dan E.B. Setiawan, “Hoax detection on social media with convolutional neural network (CNN) and support vector machine (SVM),” dalam *2023 11th Int. Conf. Inf. Commun. Technol. (ICoICT)*, 2023, hal. 361–366, doi: 10.1109/ICoICT58202.2023.10262433.

[12] X. Zhang, J. Zhao, dan Y. LeCun, “Character-level convolutional networks for text classification,” dalam *NIPS'15, Proc. 29th Int. Conf. Neural Inf. Process. Syst.*, 2015, hal. 649–657.

[13] O.C.R. Rachmawati dan Z.M.E. Darmawan, “The comparison of deep learning models for Indonesian political hoax news detection,” *CommIT (Commun. Inf. Technol.) J.*, vol. 18, no. 2, hal. 123–135, Okt. 2024, doi: 10.21512/commit.v18i2.10929.

[14] K.S. Nugroho, “Dasar text preprocessing dengan Python.” Medium. Tanggal akses: 20-Okt-2024. [Online]. Tersedia: https://ksnugroho.medium.com/dasar-text-preprocessing-dengan-python-a4fa52608ffe

[15] L. Hu, S. Wei, Z. Zhao, dan B. Wu, “Deep learning for fake news detection: A comprehensive survey,” *AI Open*, vol. 3, hal. 133–155, Okt. 2022, doi: 10.1016/j.aiopen.2022.09.001.

[16] J. Brownlee, “A gentle introduction to k-fold cross-validation.” Tanggal akses: 30-Okt-2024. [Online]. Tersedia: https://machinelearningmastery.com/k-fold-cross-validation/

[17] S. Narkhede, “Understanding confusion matrix.” Towards Datascience. Tanggal akses: 20-Okt-2024. [Online]. Tersedia: https://towardsdatascience.com/understanding-confusion-matrix-a9ad42dcfd62

[18] S.F.N. Azizah, H.D. Cahyono, S.W. Sihwi, dan W. Widiarto, “Performance analysis of transformer based models (BERT, ALBERT and RoBERTa) in fake news detection,” dalam *2023 6th Int. Conf. Inf. Commun. Technol. (ICOIACT)*, 2023, hal. 425–430, doi: 10.1109/ICOIACT59844.2023.10455849.

[19] M.A. Rosid dkk., “Improving text preprocessing for student complaint document classification using Sastrawi,” dalam *Int. Conf. Eng. Technol. Appl. Sci. (ICETsAS)*, 2020, hal. 1–6, doi: 10.1088/1757-899X/874/1/012017.

[20] M.G. Pradana, “Penggunaan fitur wordcloud dan document term matrix dalam text mining,” *J. Ilm. Inform.*, vol. 8, no. 1, hal. 38-43, Mar. 2020, doi: 10.33884/jif.v8i01.1838.

[21] R. Egger, “Text representations and word embeddings,” dalam *Applied Data Science in Tourism*. Cham, Swiss: Springer, 2022, hal. 335–361.

[22] C. Garbin, X. Zhu, dan O. Marques, “Dropout vs. batch normalization: An empirical study of their impact to deep learning,” *Multimed. Tools Appl.*, vol. 79, no. 19–20, hal. 12777–12815, Mei 2020, doi: 10.1007/s11042-019-08453-9.

[23] H. Kim dan Y.S. Jeong, “Sentiment classification using convolutional neural networks,” *Appl. Sci.*, vol. 9, no. 11, hal. 1–14, Jun. 2019, doi: 10.3390/app9112347.

[24] D.M Merkulov dan I.V. Oseledets, “Empirical study of extreme overfitting points of neural networks,” vol. 64, no. 12, hal. 1527-1534, Des. 2019, doi: 10.1134/s1064226919120118.