

Artikel

exBAKE: Model Deteksi Berita Palsu Otomatis Berdasarkan Representasi Encoder Dua Arah dari Transformers (BERT)

Heejeung Jwa ¹, Dongsuk Oh ², Kinam Park ³, Jang Mook Kang ⁴ dan Heuiseok Lim ^{1,*}

¹ Departemen Ilmu dan Teknik Komputer, Universitas Korea, Seoul 02841, Korea;

jwaheejeung@korea.ac.kr Pusat Penelitian Komputasi & AI Terinspirasi Manusia,

² Universitas Korea, Seoul 02841, Korea; inow3555@gmail.com Institut Informasi &

Komputer Kreatif, Universitas Korea, Seoul 02841, Korea; spknn@korea.ac.kr

³ Departemen Konten Konvergensi, Global Cyber University, Seoul 02841, Korea; honukang@gmail.com

⁴

* Korespondensi: limhseok@korea.ac.kr; Telp.: +82-2-3290-2396

Diterima: 4 September 2019; Diterima: 24 September 2019; Diterbitkan: 28 September 2019



Abstrak: Berita saat ini menyebar dengan cepat melalui internet. Karena berita palsu dirancang untuk menarik pembaca, mereka cenderung menyebar lebih cepat. Bagi sebagian besar pembaca, mendeteksi berita palsu dapat menjadi tantangan dan pembaca seperti itu biasanya akhirnya percaya bahwa berita palsu itu adalah fakta. Karena berita palsu dapat menjadi masalah sosial, diperlukan model yang secara otomatis mendeteksi berita palsu tersebut. Dalam makalah ini, kami fokus pada metode deteksi berita palsu otomatis berbasis data. Kami pertama-tama menerapkan model Bidirectional Encoder Representations from Transformers (BERT) untuk mendeteksi berita palsu dengan menganalisis hubungan antara judul dan teks isi berita. Untuk lebih meningkatkan kinerja, data berita tambahan dikumpulkan dan digunakan untuk melatih model ini sebelumnya. Kami menentukan bahwa sifat kontekstualisasi mendalam dari BERT paling cocok untuk tugas ini dan meningkatkan skor F 0,14 dibandingkan model canggih yang lebih lama.

Kata kunci: berita palsu; informasi palsu; mendeteksi berita palsu; tantangan berita palsu; klasifikasi berita palsu ; pembelajaran yang mendalam

1. Perkenalan

Informasi palsu muncul dalam berbagai bentuk, termasuk video, audio, gambar, dan teks. Selanjutnya, informasi palsu dalam bentuk teks dapat diklasifikasikan sebagai berita, layanan jejaring sosial, pidato, dokumen, dan sebagainya. Penelitian ini mengusulkan model pendeteksian berita palsu dengan fokus pada berita palsu berbasis teks. Berita palsu baru-baru ini menjadi masalah yang tersebar luas di seluruh dunia. Informasi yang dipalsukan atau dipalsukan dapat menyebar dengan cepat dan menjadi masalah jika pembaca secara sepintas tidak mendeteksi apakah informasi yang diberikan adalah berita palsu atau tidak.

Pada tahun 2015, Jaringan Pengecekan Fakta Internasional (IFCN) didirikan oleh Poynter, sebuah lembaga pendidikan media Amerika. IFCN mengamati tren pengecekan fakta dan menyediakan program pelatihan untuk pemeriksa fakta. Selain itu, berbagai upaya telah dilakukan untuk mencegah penyebaran berita palsu dengan memberikan kode prinsip yang dapat digunakan oleh organisasi pemeriksa fakta di seluruh dunia. Politifact (<https://www.politifact.com>) dan snopes (<https://www.snopes.com>) mengembangkan alat pendeteksi berita palsu untuk mengklasifikasikan tingkat berita palsu secara bertahap berdasarkan kriteria yang disajikan. Namun, alat ini memakan waktu dan mahal karena memerlukan pekerjaan manual dan penilaian. Oleh karena itu, diperlukan sebuah model yang secara otomatis mendeteksi berita palsu.

Ada beberapa tugas yang terlibat dalam mendeteksi berita palsu. Tantangan berita palsu tahap 1 (FNC-1) (www.fakenewschallenge.org) melibatkan pengklasifikasian pendirian teks tubuh dari artikel berita relatif

untuk sebuah judul. Teks isi mungkin setuju, tidak setuju, berdiskusi, atau tidak terkait dengan judul. Tantangan berita palsu Pencarian Web dan Penambangan Data (WSDM) 2019 (www.kaggle.com/c/fake-news-pair **klasifikasi-tantangan**) mendeteksi berita palsu dengan mengklasifikasikan judul artikel berita. Diberi judul artikel berita palsu A dan judul artikel berita mendatang B, peserta diminta untuk mengklasifikasikan B ke dalam salah satu dari tiga kategori [1–3]. Selain itu, ada tugas lain yang terlibat dalam pendeteksian berita palsu menggunakan data Layanan Jaringan Sosial (SNS). Tugas tantangan ini (www.clickbait-challenge.org) adalah mengembangkan pengklasifikasi yang menilai seberapa clickbaiting posting media sosial [4–6]. Terakhir, salah satu tugas dalam tantangan penelitian dan pengembangan kecerdasan buatan (www.ai-challenge.kr) di Korea melibatkan pendeteksian konten berita kontekstual yang tidak terkait dalam bodytext berita. Hal ini dilakukan untuk mencegah situasi di mana pembaca terkena informasi yang tidak diinginkan.

Namun, hasil dari tantangan dan tugas ini tidak menunjukkan kinerja yang baik dan dapat ditingkatkan. Selain itu, penyisipan kata merupakan faktor penting untuk meningkatkan kinerja model. word2vec [7] dan fastText [8] sebelumnya digunakan untuk penyisipan kata. Ini tidak menunjukkan kinerja yang baik karena mereka menggunakan nilai vektor tetap daripada nilai vektor cair untuk kata-kata. Untuk melengkapi ini, kami menggunakan penyisipan kata kontekstual. Model tipikal termasuk Embeddings from Language Models (ELMO) [9], BERT [10], dan Generative Pre-Training (GPT) [11].

ELMO menggunakan struktur bi-LSTM [12] dan pendekatan berbasis fitur yang membutuhkan penerapan banyak hyperparameter. Pendekatan berbasis fitur mencakup representasi bahasa pra-terlatih dari fitur tambahan jaringan yang melakukan tugas-tugas tertentu. Di sisi lain, BERT dan GPT menggunakan struktur transformator dan pendekatan fine-tuning untuk meminimalkan hyperparameter. Pendekatan fine-tuning melibatkan pengurangan parameter tugas tertentu sebanyak mungkin dan sedikit mengubah parameter yang telah dilatih sebelumnya dengan melatih tugas-tugas hilir. Kami mempertimbangkan faktor-faktor di atas dan menerapkan data berita palsu ke model BERT, memungkinkan kami untuk menganalisis makna artikel berita dengan lebih baik.

Dalam makalah ini, kami mengusulkan BAKE, model deteksi berita palsu otomatis yang meningkatkan BERT dengan mengurangi masalah ketidakseimbangan data. Selain itu, kami juga mengusulkan model yang menggabungkan copora berita ekstra tidak berlabel ke dalam BAKE, yang kami sebut exBAKE.

Kontribusi utama kami dirangkum sebagai berikut:

- Kami menggunakan BERT [10], yang merupakan studi pertama tentang deteksi berita palsu menggunakan dataset teks judul-tubuh. BERT menyertakan representasi bahasa pra-pelatihan yang dikembangkan oleh Google.
- Kami menyadari bahwa data tidak stabil, oleh karena itu untuk mengembangkan model BAKE untuk mengklasifikasikan data menggunakan weighted cross entropy (WCE).
- Kami menyertakan data berita CNN dan Daily Mail untuk pra-pelatihan BAKE. Jumlah data berita yang lebih besar digunakan untuk mendeteksi berita palsu secara lebih efisien.
- Terakhir, kami mengevaluasi kinerja model exBAKE yang diusulkan, dan menunjukkan bahwa kinerjanya lebih baik daripada model lain yang menggunakan data FNC-1.

Sisa dari makalah ini disusun sebagai berikut. Di Bagian 2, kami menyajikan gambaran umum tentang karya-karya terkait. Bagian 3 menyajikan data yang digunakan dalam penelitian ini. Di Bagian 4, kami menganalisis struktur model yang diusulkan, dan, di Bagian 5, kami menyajikan eksperimen dan hasilnya. Akhirnya, kami menyimpulkan makalah dan menyoroti beberapa arah penelitian masa depan di Bagian 6.

2. Pekerjaan Terkait

Deteksi berita palsu telah diteliti dengan menggunakan beberapa metode sesuai dengan cakupan dan format data berita palsu yang tersedia dan pendekatan teknis [13–15]. Tugas yang harus dilakukan pada kumpulan data berita palsu termasuk memverifikasi apakah judul cocok dengan teks isi, menemukan kalimat yang tidak cocok dalam teks isi, dan mengidentifikasi penyebaran berita palsu melalui SNS. Dalam penelitian ini, kami menggunakan kumpulan data untuk tugas pertama. Metode tersebut umumnya menggunakan teknik seperti pembelajaran mendalam [16,17], pembelajaran mesin [18,19], atau metode berbasis aturan [20] untuk tujuan deteksi. Dalam penelitian ini, kami menggunakan pembelajaran mendalam untuk melatih data.

Suara mayoritas mencakup fitur dasar FNC-1. Kemunculan bersama (COOC) karakter n-gram dan kata dari dokumen, judul dan dua fitur berbasis leksikon (yaitu, kata polaritas (POLA) dan

hitung jumlah sanggahan (REFU) berdasarkan daftar kata-kata kecil yang menggunakan baseline peningkatan gradien dari mana penyelenggara FNC-1 disediakan. Dengan mengamati skor FNC dari kinerja sistem, ditentukan bahwa fitur berbasis leksikon dan baseline suara mayoritas beroperasi secara sebanding. Di sisi lain, COOC menunjukkan kinerja yang relatif lebih baik.

Dalam tantangan berita palsu, tempat pertama diamankan oleh tim SWEN yang menggunakan TalosComb. Model TalosComb adalah model tertimbang rata-rata dari TalosCNN dan TalosTree [21]. Talostree didasarkan pada model pohon keputusan yang didorong oleh gradien, yang terdiri dari dekomposisi nilai tunggal (SVD) [22], jumlah kata, frekuensi dokumen terbalik frekuensi (TF-IDF) [23], dan fitur sentimen menggunakan penyematan word2vec. TalosCNN didasarkan pada jaringan saraf convolutional yang dalam dan menggunakan embeddings word2vec yang telah dilatih sebelumnya. Ini menggunakan beberapa lapisan konvolusi yang terdiri dari tiga lapisan yang terhubung penuh dan lapisan softmax akhir untuk klasifikasi.

Dalam tantangan berita palsu, tempat kedua diamankan oleh tim Athena yang mengusulkan multi-layer perceptron (MLP) [24]. Ini memperluas struktur model asli ke enam lapisan tersembunyi dan satu lapisan softmax dan menggabungkan beberapa fitur yang direkayasa secara manual dari tim Athena—yaitu, unigram, kesamaan kosinus penyisipan kata kata kerja dan kata benda antara token dokumen dan judul, alokasi Dirichlet laten, topik model berdasarkan faktorisasi matriks non-negatif, dan pengindeksan semantik laten. Selain itu, fitur dasar disediakan oleh penyelenggara FNC-1.

Jenis fitur ini baik membentuk vektor fitur terpisah atau vektor fitur gabungan [25]. Dari hasil yang diperoleh dari dataset uji FNC-1, ditentukan bahwa featMLP menunjukkan kinerja keseluruhan yang baik tetapi masih belum yang terbaik. Seperti sistem lainnya, ada banyak perbedaan antara kinerja featMLP selama pengembangan dan pada kumpulan data pengujian. Karena 100 topik baru, yang belum disertakan dalam dataset pelatihan, model stackLSTM digabungkan dengan jaringan memori jangka pendek (LSTM) [26] yang ditumpuk dan set fitur terbaik diturunkan dari tes ablasi [27]. Kelas DSC melakukan yang terbaik dalam hal stackLSTM. Ini adalah fitur penting dari stackLSTM. Karena ada beberapa contoh kelas DSG, DSG sulit untuk diklasifikasikan. Dengan kata lain, stackLSTM dengan benar mendeteksi instance negasi yang lebih kompleks.

Dalam tantangan berita palsu, tempat ketiga diamankan oleh tim UCL Machine Reading (UCLMR). Tim menyarankan MLP menggunakan satu lapisan tersembunyi [28]. Tim menggunakan frekuensi istilah (TF) dan TF-IDF untuk mengekspresikan input teks. Vektor TF diekstraksi dari kosakata 5000 kata yang paling sering muncul di set pelatihan, dan vektor TF-IDF diperoleh dari kosakata 5000 kata yang paling sering muncul baik di set data pelatihan maupun uji.

Hasil penggunaan BERT dalam makalah ini adalah kinerja yang lebih tinggi dari model yang ada. Karena data berita terdiri dari berbagai kata dan kalimat, penting untuk memahami dengan jelas hubungan antar kata untuk analisis yang akurat. BERT dirancang untuk secara jelas mengidentifikasi hubungan antara kata-kata dalam sebuah kalimat. BERT mengadopsi pembelajaran semi-diawasi dan model representasi bahasa yang hanya menggunakan bagian encoder dari transformator [29]. Secara khusus, BERT didasarkan pada encoder transformator dua arah multi-lapisan yang secara bersama-sama mengkondisikan konteks kiri dan kanan di semua lapisan. BERT melakukan pra-pelatihan menggunakan tugas prediksi tanpa pengawasan, yang mencakup model bahasa bertopeng (MLM) dan prediktor kalimat berikutnya. MLM adalah tentang memahami konteks terlebih dahulu dan kemudian memprediksi kata-kata. Pertama, kami secara acak menutupi beberapa token dengan probabilitas 15% dari input yang diterapkan potongan kata. Input disertakan dalam struktur Transformer untuk memprediksi kata-kata bertopeng berdasarkan konteks kata-kata di sekitarnya. Melalui proses ini, BERT memahami konteks lebih akurat. Prediktor kalimat selanjutnya adalah untuk mengidentifikasi hubungan antar kalimat. Tugas ini penting untuk tugas pemahaman bahasa seperti Question Answering (QA) atau Natural Language Inference (NLI). BERT menyertakan tugas prediksi kalimat biner berikutnya, yang menggabungkan dua kalimat dalam korpus dengan kalimat aslinya. Struktur model ini memungkinkan BERT untuk tampil sangat baik dalam berbagai tugas NLP. Kami menggunakan model dasar BERT, yang mencakup model dasar dan model besar, dan, tergantung pada ukuran model, kami menggunakan jumlah lapisan yang berbeda untuk blok transformator, ukuran tersembunyi, dan kepala self-attention. Data yang digunakan dalam model BERT terdiri dari 800 M kata dari Book Corpus dan 2500 M kata dari Wikipedia.

3. Data

Dalam penelitian ini, kami memasukkan CNN (www.cnn.com) dan Surat Harian (www.dailymail.co.uk) kumpulan data (<https://github.com/abisee/cnn-dailymail>) untuk pembelajaran tambahan selama tahap pra-pelatihan BERT untuk meningkatkan kemampuan deteksinya. Data yang digunakan proses pelatihan dan pengujian digunakan untuk menyempurnakan artikel berita. Pra-pelatihan BERT menunjukkan kinerja yang baik dalam tugas pemrosesan bahasa alami (NLP) sebelumnya [30–32]. Namun, data yang digunakan dalam model BERT didasarkan pada 2500 M kata dari data umum yang diperoleh dari Wikipedia dan 800 M kata dari Book Corpus.

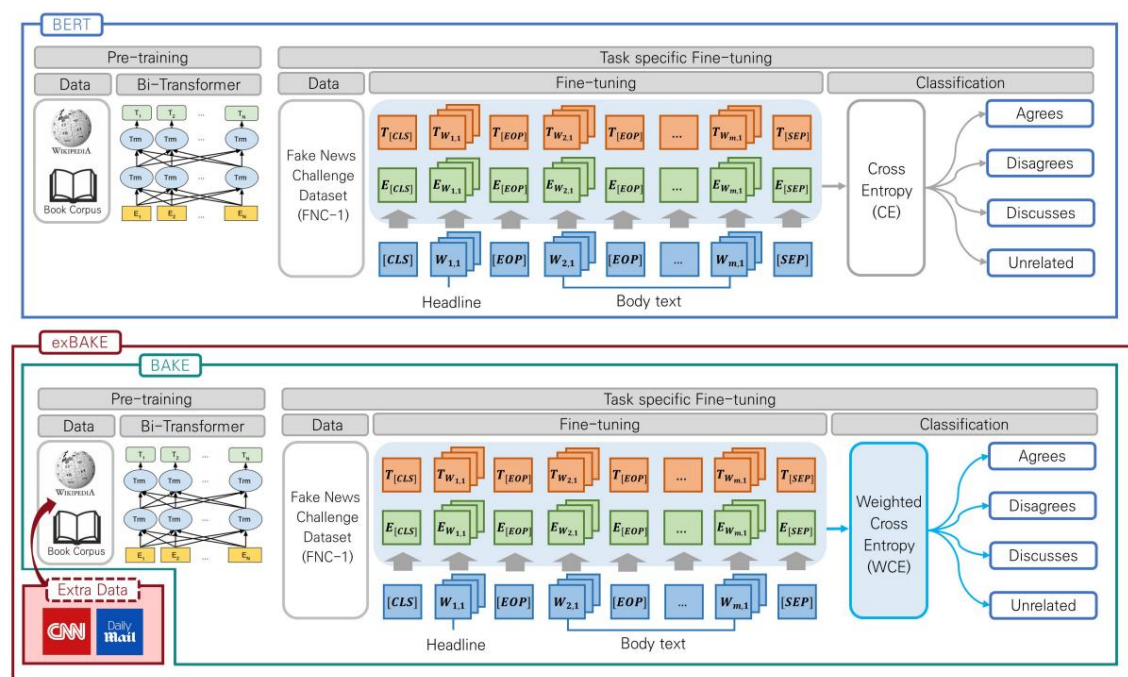
Meskipun data ini mencakup berbagai informasi, masih kekurangan informasi rinci tentang domain individu. Untuk mengatasi kekurangan ini, data berita telah ditambahkan pada tingkat pra-pelatihan dalam penelitian ini untuk meningkatkan kemampuan pendeteksian berita palsu. Data ringkasan [33–36] dari CNN mencakup sekitar 90.000 dokumen dan 380.000 pertanyaan (dengan ukuran kosakata 118.497), sedangkan kumpulan data Daily Mail mencakup 197.000 dokumen dan 879.000 pertanyaan (dengan ukuran kosakata 208.045). Artikel CNN dikumpulkan dari periode antara April 2007 hingga akhir April 2015 dari situs web CNN. Artikel Daily Mail dikumpulkan dari periode antara Juni 2010 dan akhir April 2015 dari situs Daily Mail [37].

Data FNC-1 digunakan untuk fine-tuning. Set pelatihan terdiri dari pasangan judul dan teks isi yang sesuai, termasuk label kelas yang sesuai untuk setiap pasangan. Selain itu, set tes terdiri dari pasangan judul dan teks isi tanpa label kelas untuk membantu mengevaluasi sistem. Secara total, 2587 tajuk utama dan 2587 teks isi digunakan, dan datanya dapat ditemukan di github FNC-1 (<https://github.com/FakeNewsChallenge/fnc-1>).

4. Metode

Model yang kami usulkan disajikan pada Gambar 1. Ini terutama terdiri dari dua bagian.

Dalam proses fine-tuning, kami menggunakan WCE [38–40] untuk mengklasifikasikan dataset menjadi empat kelompok: Setuju (AGR), Tidak Setuju (DSG), Diskusikan (DSC), dan Tidak Terkait (UNR). Meskipun ini pada dasarnya adalah model BERT, kami menyebutnya BAKE karena kami pertama kali menerapkannya pada tugas deteksi berita palsu dalam kasus kami. Dalam proses pra-pelatihan, kami bereksperimen dengan ekstra data berita CNN dan Daily Mail ke model BAKE kami, menciptakan model exBAKE.



Gambar 1. Ini adalah gambar yang menggambarkan BAKE dan exBAKE yang kami ajukan.

Dengan kata lain, data diklasifikasikan menjadi empat multi-kelas menggunakan lapisan Linear dan Softmax, dan WCE digunakan sebagai kehilangan pelatihan. Kehilangan pelatihan ini, seperti yang ditunjukkan di bawah ini, dicirikan oleh nilai bobot yang berbeda menurut statistik korpus dan distribusi label yang dianalisis untuk setiap kelas. Dataset FNC-1 tidak seimbang dengan AGR 7.4%, DSG 2.0%, DSC 17.7%, dan UNR 72.8%. Cross Entropy (CE) [41,42] adalah fungsi kerugian yang paling banyak digunakan. Ini adalah metode untuk menghitung jumlah informasi yang ada antara dua distribusi probabilitas, probabilitas sebenarnya P dan probabilitas yang diprediksi Q :

$$LWCE = \frac{1}{N} \sum_{i=1}^n w_i \cdot LCE_i, \quad (1)$$

$$w_i = \frac{\sum_{i=1}^n x_i}{x_i}, \quad (2)$$

$$LCE_i = [y_i \log(p_i) + (1 - y_i) \log(1 - p_i)]. \quad (3)$$

Kami menyatakan kalimat berita yang berisi judul dan teks isi dengan $s_i = w_i, 1, \dots, w_i, l_i$ dan menyatakan input BERT dengan $x = ([CLS], s_1, [EOP], s_2, [EOP], \dots, [SEP])$. Seperti pada input, kalimat berita, termasuk headline dan body text, dibagi berdasarkan tag (EOP) (yaitu, End Of Paragraph).

5. Percobaan

5.1. Metode evaluasi

Penyelenggara FNC mengusulkan metrik evaluasi hierarki FNC, tetapi mereka tidak mempertimbangkan fakta bahwa kumpulan data FNC-1 sangat tidak seimbang. Mencapai skor tinggi di FNC tidak sulit, karena hanya melakukan tugas mayoritas (UNR) dengan baik dan memprediksi yang lain secara acak masih akan menghasilkan skor yang baik. Oleh karena itu, metrik evaluasi hierarki FNC tidak sesuai untuk memvalidasi tugas deteksi posisi tingkat dokumen [25].

Dalam penelitian ini, kami menggunakan metode evaluasi skor F1 (F1) rata-rata makro [43]. F1 dapat diartikan sebagai rata-rata tertimbang presisi dan recall. F1 menghitung metrik untuk setiap label dan memperoleh rata-rata tidak berbobot. Itu diperoleh dengan menggunakan rumus yang diberikan di bawah ini:

$$\text{presisi} = \frac{\text{Presisi1} + \text{Presisi2}}{2}, \quad (4)$$

$$\text{Ingat} = \frac{\text{Ingat1} + \text{Ingat2}}{2}, \quad (5)$$

$$F1 = 2 \cdot \frac{\text{Presisi} \cdot \text{Ingat}}{\text{Presisi} + \text{Ingat}}. \quad (6)$$

5.2. Model Perbandingan

Hasilnya dilaporkan dalam Tabel 1. Kami membuktikan bahwa BERT paling cocok untuk tugas ini karena sifatnya yang mengontekstualisasikan yang dalam. Kami telah mengungguli model kami sebelumnya dengan menerapkan BERT, dan lebih meningkatkan kinerja BAKE dengan menggunakan WCE, dan exBAKE adalah hasil paling mutakhir dari mempelajari lebih banyak data berita.

Kami membandingkan kinerja metode kami dengan pendekatan sebelumnya pada Tabel 1. BAKE dan exBAKE melampaui kinerja state-of-the-art (stackLSTM) sebelumnya masing-masing sebesar 0,125 dan 0,137 skor F1. Metode yang diusulkan juga melampaui BERT, yang sudah menjadi dasar yang kuat. Ini menunjukkan bahwa penggunaan WCE sangat penting untuk menunjukkan hasil yang kompetitif dalam deteksi berita palsu. Model exBAKE telah menunjukkan skor F1 keseluruhan terbaik, yang menunjukkan bahwa menggabungkan pengetahuan ekstra dari korpora berita besar bermanfaat untuk tugas ini.

Tabel 1. Performa model. Kami meningkatkan skor F 0,14 dibandingkan hasil canggi sebelumnya.

model	F1 AGR DSG DSC UNR
Suara mayoritas	0,210 0,0 0,0 0,0 0,839
TalosComb [21]	0,582 0,539 0,035 0,760 0,994
Pohon Talos [21]	0,570 0,520 0,003 0,762 0,994
TalosCNN (21)	0,308 0,258 0,092 0,0 0,882
Athena [25]	0,604 0,487 0,151 0,780 0,996
UCLMR [28]	0,583 0,479 0,114 0,747 0,899
featMLP [24,25]	0,607 0,530 0,151 0,766 0,982
stackLSTM [25,27]	0,609 0,501 0,180 0,757 0,995
BERT	0,656 0,651 0,145 0,839 0,989
MEMANGGANG	0,734 0,667 0,463 0,822 0,986
exBAKE	0,746 0,684 0,501 0,813 0,988
Batas atas	0,754 0,588 0,667 0,765 0,997

Secara keseluruhan, metode kami mampu mencapai kinerja mutakhir dalam tiga dari empat berita kategori. Selanjutnya, metode yang diusulkan melampaui batas atas (skor manusia annotator), dalam AGR dan DSC. Perbedaan kinerja paling dramatis dalam kategori minoritas, kembali menunjukkan bahwa WCE berperan penting dalam mengatasi masalah ketidakseimbangan data. Ini datang dengan harga penurunan kecil dalam kinerja di kategori mayoritas. Tetap saja, kinerjanya pada kategori mayoritas ini lebih unggul atau sebanding dengan state-of-the-art sebelumnya.

6. Kesimpulan

Sebagian besar data yang dikumpulkan untuk deteksi berita palsu ditulis dalam bahasa Inggris. Sebagai penyebaran berita palsu memberikan dampak negatif bagi masyarakat, beberapa penelitian telah dilakukan, dan banyak teknologi telah diperkenalkan untuk menangani teks-teks yang dipalsukan tersebut. Sangat penting untuk memungkinkan pembaca membedakan antara berita asli dan berita palsu.

Dalam penelitian ini, kami mengusulkan model exBAKE yang ditingkatkan dengan menggunakan pra-pelatihan berdasarkan BERT model untuk secara akurat memahami isi artikel tersebut. Hasilnya menunjukkan bahwa model bekerja paling baik pada kumpulan data FNC-1, yang mendeteksi berita palsu dengan menganalisis hubungan antara headline dan teks tubuh yang sesuai dari artikel berita.

Tidak ada alat otomatis yang sebelumnya dibuat untuk memeriksa keaslian artikel berita secara nyata waktu. Model yang kami usulkan akan membantu pembaca dan jurnalis lain untuk menghindari keharusan membuka secara manual melalui proses membedakan berita palsu dari berita nyata.

Di masa depan, kami akan bereksperimen dengan berbagai kasus tugas pendeteksian berita palsu menggunakan model BERT pra-terlatih diusulkan dalam penelitian ini. Kami hanya menganalisis hubungan antara judul dan isi teks artikel. Eksperimen lebih lanjut diperlukan untuk menerapkan data dari tugas pendeteksian berita palsu ke model BERT, yang akan menggunakan data berita tambahan pada fase pra-pelatihan.

Kontribusi Penulis: Konseptualisasi, kurasi data, analisis formal, investigasi, metodologi, perangkat lunak, visualisasi, penulisan—persiapan draft asli, dan penulisan—review dan editing, HJ dan DO; validasi, pengawasan, sumber daya, administrasi proyek dan perolehan dana, KP, JMK dan HL

Ucapan Terima Kasih: Karya ini didukung oleh Institut Teknologi Informasi & Komunikasi Hibah Promosi (IITP) yang didanai oleh pemerintah Korea (MSIT) (IITP-2018-0-00705, Desain dan perangkat lunak algoritma pemodelan untuk menilai berita palsu berdasarkan Artificial Intelligence. Karya ini didukung oleh National Research Hibah Foundation of Korea (NRF) yang didanai oleh pemerintah Korea (MSIP) (No. NRF-2017M3C4A7068189).

Konflik Kepentingan: Para penulis menyatakan tidak ada konflik kepentingan.

Referensi

1. Pham, L. Mentransfer, Mengubah, Ensembling: Rumus Novel Mengidentifikasi Berita Palsu. Dalam Proceedings of the 12th ACM International Conference on Web Search and Data Mining, Melbourne, Australia, 11–15 Februari 2019.
2. Liu, S.; Liu, S.; Ren, L. Percaya atau Tersangka? Kerangka Ensemble Empiris untuk Klasifikasi Berita Palsu. Dalam Proceedings of the 12th ACM International Conference on Web Search and Data Mining, Melbourne, Australia, 11–15 Februari 2019.
3. Yang, KC; Niven, T.; Kao, HY Deteksi Berita Palsu sebagai Inferensi Bahasa Alami. Dalam Proceedings of the 12th ACM International Conference on Web Search and Data Mining, Melbourne, Australia, 11–15 Februari 2019.
4. Omidvar, A.; Jiang, H.; An, A. Menggunakan Neural Network untuk Mengidentifikasi Clickbaits di Media Berita Online. Dalam Simposium Internasional Tahunan tentang Manajemen Informasi dan Data Besar, Lima, Peru, 3-5 September 2018; Pegas: Cham, Swiss, 2018; hal.220–232.
5. Zhou, Y. Deteksi clickbait dalam tweet menggunakan jaringan self-attentive. arXiv **2017**, arXiv:1710.05364.
6. Grigorev, A. Mengidentifikasi postingan clickbait di media sosial dengan ansambel model linier. arXiv **2017**, arXiv: 1710.00399.
7. Mikolov, T.; Sutskever, I.; Chen, K.; Corrado, GS; Dean, J. Representasi terdistribusi dari kata dan frasa dan komposisinya. Dalam Prosiding Konferensi Internasional ke-26 tentang Sistem Pemrosesan Informasi Saraf, Lake Tahoe, NV, AS, 5-10 Desember 2013; hal. 3111–3119.
8. Bojanowski, P.; Kuburan, E.; Joulin, A.; Mikolov, T. Memperkaya vektor kata dengan informasi subkata. Trans. Asosiasi Hitung. Ahli bahasa. **2017**, 5, 135–146. [CrossRef]
9. Peters, S.; Neumann, M.; Iyyer, M.; Gardner, M.; Clark, C.; Lee, K.; Zettlemoyer, L. Representasi kata yang dikontekstualisasikan secara mendalam. arXiv **2018**, arXiv:1802.05365.
10. Devlin, J.; Chang, MW; Lee, K.; Toutanova, K. Bert: Pra-pelatihan transformator dua arah dalam untuk pemahaman bahasa. arXiv **2018**, arXiv:1810.04805.
11. Radford, A.; Narasimhan, K.; Salimán, T.; Sutskever, I. Meningkatkan Pemahaman Bahasa dengan Pra-Pelatihan Generatif. 2018. Tersedia online: [https://s3-us-west-2.amazonaws.com/openai-assets/researchcovers/ languageunsupervised/bahasapemahamanpaper.pdf](https://s3-us-west-2.amazonaws.com/openai-assets/researchcovers/languageunsupervised/bahasapemahamanpaper.pdf) (diakses pada 20 September 2019).
12. Schuster, M.; Paliwal, KK Jaringan saraf berulang dua arah. IEEE Trans. Proses Sinyal. **1997**, 45, 2673–2681. [CrossRef]
13. Wang, WY “Pembohong, celana pembohong terbakar”: Dataset benchmark baru untuk deteksi berita palsu. arXiv **2017**, arXiv:1705.00648.
14. Ruchansky, N.; Seo, S.; Liu, Y. Csi: Model mendalam hibrida untuk pendeteksian berita palsu. In Proceedings of the 2017 ACM on Conference on Information and Knowledge Management, Singapura, 6–10 November 2017; ACM: New York, NY, AS, 2017; hlm. 797–806.
15. Kochkina, E.; Liakata, M.; Augenstein, I. Turing di semester-2017 tugas 8: Pendekatan berurutan untuk klasifikasi posisi rumor dengan branch-lstm. arXiv **2017**, arXiv:1704.07221.
16. Popat, K.; Mukherjee, S.; Yates, A.; Weikum, G. DeClarE: Membongkar berita palsu dan klaim palsu menggunakan pembelajaran mendalam yang sadar akan bukti. arXiv **2018**, arXiv:1809.06416.
17. Yang, Y.; Zheng, L.; Zhang, J.; Cui, Q.; Li, Z.; Yu, PS TI-CNN: Jaringan saraf convolutional untuk berita palsu deteksi. arXiv **2018**, arXiv:1806.00749.
18. Rasool, T.; pantat, WH; Shaukat, A.; Akram, MU Multi-Label Deteksi Berita Palsu menggunakan Multi-layered Supervised Learning. Dalam Prosiding Konferensi Internasional ke-11 2019 tentang Teknik Komputer dan Otomasi , Perth, Australia, 23–25 Februari 2019; ACM: New York, NY, AS, 2019; hlm. 73–77.
19. Yang, S.; Shu, K.; Wang, S.; Gu, R.; Wu, F.; Liu, H. Deteksi berita palsu tanpa pengawasan di media sosial: Pendekatan generatif. Dalam Prosiding Konferensi AAAI ke-33 tentang Kecerdasan Buatan, Honolulu, HI, AS, 27 Januari-1 Februari 2019.
20. Feng, S.; Banerjee, R.; Choi, Y. Stylometry sintaksis untuk deteksi penipuan. Dalam Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Short Papers—Volume 2, Pulau Jeju, Korea, 8–14 Juli 2012; hal.171–175.

21. Sean, B.; Doug, S.; Yuxi, P. Talos Targetkan Disinformasi dengan Kemenangan Tantangan Berita Palsu. 2017. Tersedia online: <https://blog.talosintelligence.com/2017/06/talos-fake-news-challenge.html> (diakses pada 20 September 2019).
22. De Lathauwer, L.; De Moor, B.; Vandewalle, J. Dekomposisi nilai singular multilinear. *SIAM J. Matriks anal aplikasi* **2000**, 21, 1253–1278. [[CrossRef](#)]
23. Ramos, J. Menggunakan tf-idf untuk menentukan relevansi kata dalam query dokumen. Dalam *Prosiding Konferensi Instruksional Pertama tentang Pembelajaran Mesin*, Piscataway, NJ, AS, 3–8 Desember 2003; Jilid 242, hal.133-142.
24. Davis, R.; Proctor, C. Berita palsu, konsekuensi nyata: Merekrut jaringan saraf untuk memerangi berita palsu. 2017. Tersedia online: <https://web.stanford.edu/class/archive/cs/cs224n/cs224n.1174/reports/2761239.pdf> (diakses pada 20 September 2019).
25. Hanselowski, A.; PVS, A.; Schiller, B.; Caspelherr, F.; Chaudhuri, D.; Meyer, CM; Gurevych, I. Analisis retrospektif dari tugas deteksi sikap tantangan berita palsu. *arXiv* **2018**, arXiv:1806.05180.
26. Hochreiter, S.; Schmidhuber, J. Memori jangka pendek panjang. *Komputer Saraf*. **1997**, 9, 1735–1780. [[CrossRef](#)] [[PubMed](#)]
27. Hermans, M.; Schrauwen, B. Pelatihan dan Analisis Jaringan Syaraf Tiruan Dalam. *Kemajuan dalam Sistem Pemrosesan Informasi Saraf 26*; Burges, CJC, Bottou, L., Welling, M., Ghahramani, Z., Weinberger, KQ, Eds.; Curran Associates, Inc.: New York, NY, AS, 2013; hal 190–198.
28. Riedel, B.; Augenstein, saya.; Spithourakis, GP; Riedel, S. Garis dasar yang sederhana namun sulit dikalahkan untuk tugas pendeteksian sikap Tantangan Berita Palsu. *arXiv* **2017**, arXiv:1707.03264.
29. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, AN; Kaiser, .; Polosukhin, I. Perhatian adalah semua yang Anda butuhkan. Dalam *Prosiding Konferensi ke-31 tentang Sistem Pemrosesan Informasi Saraf*, Long Beach, CA, AS, 4–9 Desember 2017; hal.5998–6008.
30. Diaz, M.; Ferrer, MA; Impedovo, D.; Pirlo, G.; Vessio, G. Representasi tulisan tangan statis yang ditingkatkan secara dinamis untuk deteksi penyakit Parkinson. *Pengenalan Pola. Lett.* **2019**, 128, 204–210. [[CrossRef](#)]
31. Lee, J.; Yoon, W.; Kim, S.; Kim, D.; Kim, S.; Jadi, CH; Kang, J. Biobert: Model representasi bahasa biomedis terlatih untuk penambangan teks biomedis. *arXiv* **2019**, arXiv:1901.08746.
32. Liu, Y.; Lapata, M. Peringkasan Teks dengan Encoder Terlatih. *arXiv* **2019**, arXiv:1908.08345.
33. Lihat, A.; Liu, PJ; Manning, CD Langsung ke intinya: Ringkasan dengan jaringan pointer-generator. *arXiv* **2017**, arXiv:1704.04368.
34. Liu, Y. Sempurnakan BERT untuk Peringkasan Ekstraktif. *arXiv* **2019**, arXiv:1903.10318.
35. Liu, L.; Lu, Y.; Yang, M.; Ku, T.; Zhu, J.; Li, H. Jaringan permusuhan generatif untuk peringkasan teks abstrak. Dalam *Prosiding Konferensi AAAI Tiga Puluh Dua tentang Kecerdasan Buatan*, New Orleans, LA, AS, 2–7 Februari 2018.
36. Paulus, R.; Xiong, C.; Socher, R. Model yang diperkuat dalam untuk peringkasan abstrak. *arXiv* **2017**, arXiv:1705.04304.
37. Hermann, KM; Kocisky, T.; Grefenstette, E.; Espeholt, L.; Kay, W.; Sulaiman, M.; Blunsom, P. Mesin Pengajaran untuk Membaca dan Memahami. *Kemajuan dalam Sistem Pemrosesan Informasi Saraf 28*; Cortes, C., Lawrence, ND, Lee, DD, Sugiyama, M., Garnett, R., Eds.; Curran Associates, Inc.: New York, NY, AS, 2015; hlm. 1693–1701.
38. Yang, K.; Lee, D.; Apa, T.; Lee, S.; Lim, H. EmotionX-KU: Pengklasifikasi Emosi Kontekstual berbasis BERT-Max. *arXiv* **2019**, arXiv:1906.11565.
39. Aurelio, YS; de Almeida, GM; de Castro, CL; Braga, AP Belajar dari set data yang tidak seimbang dengan fungsi cross-entropy berbobot. *Proses saraf. Lett.* **2019**, 1–13, doi:10.1007/s11063-018-09977-1. [[CrossRef](#)]
40. Sudre, CH; Li, W.; Vercauteren, T.; Ourselin, S.; Cardoso, MJ Generalized Dadu Tumpang Tindih sebagai Fungsi Kehilangan Pembelajaran Mendalam untuk Segmentasi yang Sangat Tidak Seimbang. Dalam *Pembelajaran Mendalam dalam Analisis Gambar Medis dan Pembelajaran Multimodal untuk Pendukung Keputusan Klinis*; Musim Semi: New York, NY, AS, 2017; hal.240–248.
41. Neyshabur, B.; Bhojanapalli, S.; McAllester, D.; Srebro, N. Menggali generalisasi dalam deep learning. Dalam *Prosiding Konferensi Internasional ke-31 tentang Sistem Pemrosesan Informasi Saraf*, Long Beach, CA, AS, 4–9 Desember 2017; hal.5947–5956.

42. Zhang, Z.; Sabuncu, M. Kehilangan entropi lintas umum untuk melatih jaringan saraf dalam dengan label bising. Dalam Prosiding Konferensi Internasional ke-32 tentang Sistem Pemrosesan Informasi Saraf, Montreal, QC, Kanada, 3–8 Desember 2018; hal.8778–8788.
43. Shang, W.; Huang, H.; Zhu, H.; Lin, Y.; Qu, Y.; Wang, Z. Algoritme pemilihan fitur baru untuk teks kategorisasi. Sistem Ahli. aplikasi **2007**, *33*, 1-5. [[CrossRef](#)]



© 2019 oleh penulis. Penerima Lisensi MDPI, Basel, Swiss. Artikel ini adalah artikel akses terbuka yang didistribusikan di bawah syarat dan ketentuan Creative Commons Attribution (CC BY) lisensi (<http://creativecommons.org/licenses/by/4.0/>).