

Deteksi Berita Palsu Otomatis Berbahasa Indonesia Menggunakan BERT

Reza Fuad Rachmadi
Departemen Teknik Komputer
Fakultas Teknologi Elektro
dan Informatika Cerdas
Institut Teknologi Sepuluh Nopember
Surabaya, Indonesia 60111
fuad@te.its.ac.id

Mauridhi Hery Purnomo
Departemen Teknik Komputer
Fakultas Teknologi Elektro
dan Informatika Cerdas
Institut Teknologi Sepuluh Nopember
Surabaya, Indonesia 60111
hery@ee.its.ac.id

Aufa Nabil Amiri
Departemen Teknik Komputerr
Fakultas Teknologi Elektro
dan Informatika Cerdas
Institut Teknologi Sepuluh Nopember
Surabaya, Indonesia 60111
aufa.17072@mhs.its.ac.id

Abstrak—Berita palsu atau yang biasa disebut hoaks adalah suatu yang hal yang sering melanda Indonesia. Dengan adanya sosial media, suatu berita palsu dapat memiliki tingkat penyebaran yang sangat luas. Selain itu, masyarakat Indonesia memiliki tingkat kecenderungan untuk menyebarkan berita palsu yang cukup tinggi. Sehingga, suatu metode pendeteksi berita palsu harus ada. Penelitian ini memanfaatkan algoritma BERT yang digunakan untuk mendeteksi apakah suatu berita adalah berita hoaks atau tidak secara otomatis. Dari suatu teks yang mentah, akan dilakukan tokenisasi sebelum akhirnya dimasukkan ke dalam algoritma BERT. Selanjutnya, keluaran dari BERT akan dijadikan sebagai inputan dari algoritma klasifikasi Linear Regression. Barulah pada saat ini, kita bisa mendapatkan klasifikasi apakah suatu teks itu berupa berita hoaks atau tidak. Tujuan dari penelitian ini adalah untuk membuat sebuah model yang dapat digunakan untuk melakukan klasifikasi suatu teks apakah termasuk ke dalam berita hoaks atau tidak.

Kata kunci—BERT, Hoaks, Klasifikasi, Linear Regression

I. LATAR BELAKANG

Berita adalah laporan atau cerita faktual yang disajikan paling cepat, memiliki pemaparan masalah yang baik, serta berlaku adil kepada seluruh masalah yang disajikan [1]. Berita memiliki peran yang sangat penting dalam masyarakat karena sebagai media yang dapat digunakan untuk mengetahui peristiwa yang paling baru, juga dapat digunakan sebagai media untuk menambah wawasan.

Hoaks atau berita palsu adalah sebuah cara atau usaha yang berusaha untuk menipu orang sehingga mempercayai sesuai yang salah sebagai hal benar dan seringnya hal yang salah tersebut sama sekali tidak masuk akal [2]. Selain kerugian dalam hal pengetahuan, berita palsu memiliki efek yang beragam, seperti kerugian dalam bentuk reputasi, harta benda, sampai ancaman pembunuhan.

Berdasarkan data yang diperoleh dari Kementerian Komunikasi dan Informatika total jumlah berita palsu yang ditemukan pada tahun Agustus 2018 sampai dengan Maret 2020 berjumlah 5156. pada bulan Januari 2020 sampai Maret 2020, sudah terdapat 959 berita palsu yang ditemukan [3]. Masih dari sumber yang sama, pada bulan Juni 2020, hampir setiap harinya ditemukan puluhan berita palsu baru [4].

Berita hoaks juga memiliki tingkat penyebaran yang cepat seiring dengan semakin tingginya penggunaan media sosial

oleh masyarakat. Berdasarkan survey yang dilakukan oleh Khan dan Idris, lebih dari 50% masyarakat Indonesia memiliki tingkat kecenderungan untuk melakukan share suatu tautan berita tanpa melakukan validasi terlebih dahulu [5]. Survey lain yang dilakukan oleh Kunto yang melibatkan 480 responden di Kota Jawa Barat menunjukkan bahwa sekitar 30% masyarakat Jawa Barat memiliki kecenderungan menengah sampai tinggi untuk menyebarkan berita palsu [6]. Dari sampel tersebut, dapat disimpulkan bahwa Indonesia memiliki kecenderungan tinggi untuk menyebarkan berita palsu.

Neural Networks adalah salah satu cabang dalam pembelajaran mesin yang menerapkan *neurons* layaknya struktur otak manusia untuk memproses data dan menghasilkan keluaran. Salah satu metode *neural network* yang cukup baru adalah *Bi-directional Encoder Representations from Transformers* atau disingkat sebagai BERT. BERT adalah metode yang digunakan untuk mendapatkan suatu konteks dalam suatu teks yang dimasukkan.

Pembahasan pada paper ini dimulai dengan presentasi mengenai penelitian lain (Bagian II). Kemudian dilanjutkan dengan penjelasan mengenai desain dan implementasi dari sistem yang dibuat (Bagian III). Berdasarkan hal tersebut, kami menunjukkan lorem ipsum (Bagian IV). Terakhir, didapatkan kesimpulan dari penelitian yang telah dilakukan (Bagian V).

II. PENELITIAN TERKAIT

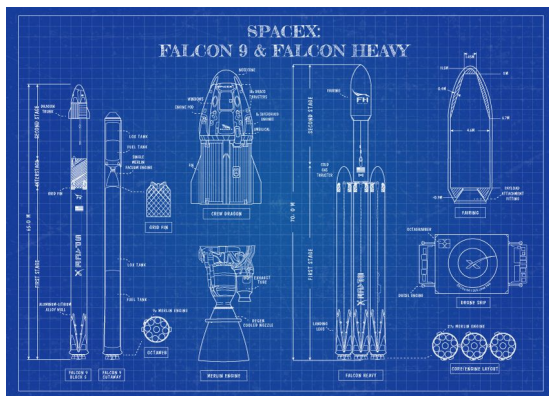
Sudah terdapat beberapa penelitian yang pernah dilakukan oleh orang lain mengenai pendeteksi berita hoaks ini. Aggarway et al. pernah melakukan penelitian untuk membandingkan antara BERT, XGBoost dan LSTM untuk melakukan klasifikasi berita palsu berbahasa Inggris. Dari penelitian tersebut didapatkan bahwa BERT memiliki tingkat akurasi yang lebih tinggi apabila dibandingkan dengan XGBoost dan LSTM [7]. Bahad et al. melakukan penelitian yang membandingkan antara CNN, RNN, *uni-directional* LSTM RNN dan *bi-directional* LSTM RNN. Hasil dari penelitian tersebut menunjukkan bahwa penggunaan LSTM ditambah dengan *attention* baik itu *uni-directional* maupun *bi-directional* memiliki tingkat akurasi yang lebih tinggi apabila dibandingkan dengan CNN atau

RNN [8]. Dari kedua penelitian tersebut, dapat diambil kesimpulan bahwa algoritma yang 'mengingat' atau mengetahui suatu konteks dalam teks akan memiliki tingkat akurasi yang lebih tinggi dibanding algoritma dengan pendekatan yang lain.

Untuk penelitian pendeteksi berita hoaks dengan berbahasa indonesia, terdapat beberapa penelitian yang pernah dilakukan seperti oleh Prasetijo et al. yang meneliti penggunaan SVM dan SGD untuk mendeteksi berita hoaks berbahasa indonesia. Penelitian tersebut berhasil membuat suatu model dengan tingkat akurasi sebesar 85% [9]. Penelitian lain yang dilakukan oleh Rahutomo et al. dengan menggunakan algoritma *naive bayes* berhasil menghasilkan akurasi sebesar 80% [10].

III. DESAIN DAN IMPLEMENTASI SISTEM

Pada cetak biru yang tertera pada Gambar 1. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Donec odio elit, dictum in, hendrerit sit amet, egestas sed, leo. Praesent feugiat sapien aliquet odio. Integer vitae justo. Aliquam vestibulum fringilla lorem. Sed neque lectus, consectetur at, consectetur sed, eleifend ac, lectus. Nulla facilisi. Pellentesque eget lectus. Proin eu metus. Sed porttitor. In hac habitasse platea dictumst. Suspendisse eu lectus. Ut mi mi, lacinia sit amet, placerat et, mollis vitae, dui. Sed ante tellus, tristique ut, iaculis eu, malesuada ac, dui. Mauris nibh leo, facilisis non, adipiscing quis, ultrices a, dui.



Gambar 1. Cetak biru roket yang akan diuji coba. [11]

Morbi luctus, wisi viverra faucibus pretium, nibh est placerat odio, nec commodo wisi enim eget quam. Quisque libero justo, consectetur a, feugiat vitae, porttitor eu, libero. Suspendisse sed mauris vitae elit sollicitudin malesuada. Maecenas ultricies eros sit amet ante. Ut venenatis velit. Maecenas sed mi eget dui varius euismod. Phasellus aliquet volutpat odio. Vestibulum ante ipsum primis in faucibus orci luctus et ultrices posuere cubilia Curae; Pellentesque sit amet pede ac sem eleifend consectetur. Nullam elementum, urna vel imperdiet sodales, elit ipsum pharetra ligula, ac pretium ante justo a nulla. Curabitur tristique arcu eu metus. Vestibulum lectus. Proin mauris. Proin eu nunc eu urna hendrerit faucibus. Aliquam auctor, pede consequat laoreet varius, eros tellus scelerisque quam, pellentesque hendrerit ipsum dolor sed augue. Nulla nec lacus.

Tabel I
CONTOH TABEL SEDERHANA

Heading1	Heading2	Heading3
One Four	Two Five	Three Six

Suspendisse vitae elit. Aliquam arcu neque, ornare in, ullamcorper quis, commodo eu, libero. Fusce sagittis erat at erat tristique mollis. Maecenas sapien libero, molestie et, lobortis in, sodales eget, dui. Morbi ultrices rutrum lorem. Nam elementum ullamcorper leo. Morbi dui. Aliquam sagittis. Nunc placerat. Pellentesque tristique sodales est. Maecenas imperdiet lacinia velit. Cras non urna. Morbi eros pede, suscipit ac, varius vel, egestas non, eros. Praesent malesuada, diam id pretium elementum, eros sem dictum tortor, vel consectetur odio sem sed wisi.

Sed feugiat. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Ut pellentesque augue sed urna. Vestibulum diam eros, fringilla et, consectetur eu, nonummy id, sapien. Nullam at lectus. In sagittis ultrices mauris. Curabitur malesuada erat sit amet massa. Fusce blandit. Aliquam erat volutpat. Aliquam euismod. Aenean vel lectus. Nunc imperdiet justo nec dolor.

```
#include <iostream>

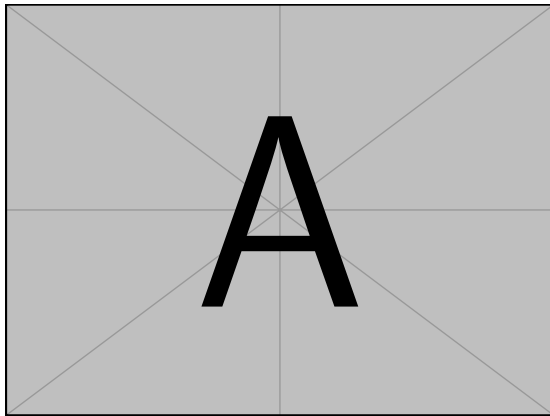
int main() {
    std::cout << "Halo_Dunia!";
    return 0;
}
```

Listing 1. Program halo dunia.

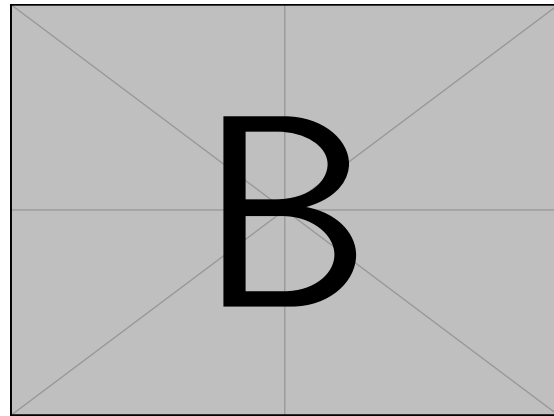
Etiam euismod. Fusce facilisis lacinia dui. Suspendisse potenti. In mi erat, cursus id, nonummy sed, ullamcorper eget, sapien. Praesent pretium, magna in eleifend egestas, pede pede pretium lorem, quis consectetur tortor sapien facilisis magna. Mauris quis magna varius nulla scelerisque imperdiet. Aliquam non quam. Aliquam porttitor quam a lacus. Praesent vel arcu ut tortor cursus volutpat. In vitae pede quis diam bibendum placerat. Fusce elementum convallis neque. Sed dolor orci, scelerisque ac, dapibus nec, ultricies ut, mi. Duis nec dui quis leo sagittis commodo.

- 1) Aliquam lectus. Vivamus leo. Quisque ornare tellus ullamcorper nulla. Mauris porttitor pharetra tortor.
- 2) Sed fringilla justo sed mauris. Mauris tellus. Sed non leo. Nullam elementum, magna in cursus sodales, augue est scelerisque sapien, venenatis congue nulla arcu et pede.
- 3) Ut suscipit enim vel sapien. Donec congue. Maecenas urna mi, suscipit in, placerat ut, vestibulum ut, massa. Fusce ultrices nulla et nisl.

Etiam ac leo a risus tristique nonummy. Donec dignissim tincidunt nulla. Vestibulum rhoncus molestie odio. Sed lobortis, justo et pretium lobortis, mauris turpis condimentum augue, nec ultricies nibh arcu pretium enim. Nunc purus neque, placerat id, imperdiet sed, pellentesque nec, nisl. Ves-



(a) Hasil A



(b) Hasil B

Gambar 2. Contoh input beberapa gambar.

tibulum imperdiet neque non sem accumsan laoreet. In hac habitasse platea dictumst. Etiam condimentum facilisis libero. Suspendisse in elit quis nisl aliquam dapibus. Pellentesque auctor sapien. Sed egestas sapien nec lectus. Pellentesque vel dui vel neque bibendum viverra. Aliquam porttitor nisl nec pede. Proin mattis libero vel turpis. Donec rutrum mauris et libero. Proin euismod porta felis. Nam lobortis, metus quis elementum commodo, nunc lectus elementum mauris, eget vulputate ligula tellus eu neque. Vivamus eu dolor.

Nulla in ipsum. Praesent eros nulla, congue vitae, euismod ut, commodo a, wisi. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Aenean nonummy magna non leo. Sed felis erat, ullamcorper in, dictum non, ultricies ut, lectus. Proin vel arcu a odio lobortis euismod. Vestibulum ante ipsum primis in faucibus orci luctus et ultrices posuere cubilia Curae; Proin ut est. Aliquam odio. Pellentesque massa turpis, cursus eu, euismod nec, tempor congue, nulla. Duis viverra gravida mauris. Cras tincidunt. Curabitur eros ligula, varius ut, pulvinar in, cursus faucibus, augue.

IV. LOREM IPSUM

Nulla mattis luctus nulla. Duis commodo velit at leo. Aliquam vulputate magna et leo. Nam vestibulum ullamcorper leo. Vestibulum condimentum rutrum mauris. Donec id mauris. Morbi molestie justo et pede. Vivamus eget turpis sed nisl cursus tempor. Curabitur mollis sapien condimentum nunc. In wisi nisl, malesuada at, dignissim sit amet, lobortis in, odio. Aenean consequat arcu a ante. Pellentesque porta elit sit amet orci. Etiam at turpis nec elit ultricies imperdiet. Nulla facilisi. In hac habitasse platea dictumst. Suspendisse viverra aliquam risus. Nullam pede justo, molestie nonummy, scelerisque eu, facilisis vel, arcu.

Curabitur tellus magna, porttitor a, commodo a, commodo in, tortor. Donec interdum. Praesent scelerisque. Maecenas posuere sodales odio. Vivamus metus lacus, varius quis, imperdiet quis, rhoncus a, turpis. Etiam ligula arcu, elementum a, venenatis quis, sollicitudin sed, metus. Donec nunc pede,

tincidunt in, venenatis vitae, faucibus vel, nibh. Pellentesque wisi. Nullam malesuada. Morbi ut tellus ut pede tincidunt porta. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam congue neque id dolor.

Donec et nisl at wisi luctus bibendum. Nam interdum tellus ac libero. Sed sem justo, laoreet vitae, fringilla at, adipiscing ut, nibh. Maecenas non sem quis tortor eleifend fermentum. Etiam id tortor ac mauris porta vulputate. Integer porta neque vitae massa. Maecenas tempus libero a libero posuere dictum. Vestibulum ante ipsum primis in faucibus orci luctus et ultrices posuere cubilia Curae; Aenean quis mauris sed elit commodo placerat. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos hymenaeos. Vivamus rhoncus tincidunt libero. Etiam elementum pretium justo. Vivamus est. Morbi a tellus eget pede tristique commodo. Nulla nisl. Vestibulum sed nisl eu sapien cursus rutrum.

```
def apakahBilanganPrima(nilai):  
    if nilai > 1:  
        for i in range(2, nilai):  
            if (nilai % i) == 0:  
                return False  
        else:  
            return True  
    else:  
        return False
```

Listing 2. Program perhitungan bilangan prima.

Nulla non mauris vitae wisi posuere convallis. Sed eu nulla nec eros scelerisque pharetra. Nullam varius. Etiam dignissim elementum metus. Vestibulum faucibus, metus sit amet mattis rhoncus, sapien dui laoreet odio, nec ultricies nibh augue a enim. Fusce in ligula. Quisque at magna et nulla commodo consequat. Proin accumsan imperdiet sem. Nunc porta. Donec feugiat mi at justo. Phasellus facilisis ipsum quis ante. In ac elit eget ipsum pharetra faucibus. Maecenas viverra nulla in massa.

Nulla ac nisl. Nullam urna nulla, ullamcorper in, interdum sit amet, gravida ut, risus. Aenean ac enim. In luctus. Phasellus

eu quam vitae turpis viverra pellentesque. Duis feugiat felis ut enim. Phasellus pharetra, sem id porttitor sodales, magna nunc aliquet nibh, nec blandit nisl mauris at pede. Suspendisse risus risus, lobortis eget, semper at, imperdiet sit amet, quam. Quisque scelerisque dapibus nibh. Nam enim. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Nunc ut metus. Ut metus justo, auctor at, ultrices eu, sagittis ut, purus. Aliquam aliquam.

V. KESIMPULAN

Etiam pede massa, dapibus vitae, rhoncus in, placerat posuere, odio. Vestibulum luctus commodo lacus. Morbi lacus dui, tempor sed, euismod eget, condimentum at, tortor. Phasellus aliquet odio ac lacus tempor faucibus. Praesent sed sem. Praesent iaculis. Cras rhoncus tellus sed justo ullamcorper sagittis. Donec quis orci. Sed ut tortor quis tellus euismod tincidunt. Suspendisse congue nisl eu elit. Aliquam tortor diam, tempus id, tristique eget, sodales vel, nulla. Praesent tellus mi, condimentum sed, viverra at, consectetur quis, lectus. In auctor vehicula orci. Sed pede sapien, euismod in, suscipit in, pharetra placerat, metus. Vivamus commodo dui non odio. Donec et felis.

Etiam suscipit aliquam arcu. Aliquam sit amet est ac purus bibendum congue. Sed in eros. Morbi non orci. Pellentesque mattis lacinia elit. Fusce molestie velit in ligula. Nullam et orci vitae nibh vulputate auctor. Aliquam eget purus. Nulla auctor wisi sed ipsum. Morbi porttitor tellus ac enim. Fusce ornare. Proin ipsum enim, tincidunt in, ornare venenatis, molestie a, augue. Donec vel pede in lacus sagittis porta. Sed hendrerit ipsum quis nisl. Suspendisse quis massa ac nibh pretium cursus. Sed sodales. Nam eu neque quis pede dignissim ornare. Maecenas eu purus ac urna tincidunt congue.

Donec et nisl id sapien blandit mattis. Aenean dictum odio sit amet risus. Morbi purus. Nulla a est sit amet purus venenatis iaculis. Vivamus viverra purus vel magna. Donec in justo sed odio malesuada dapibus. Nunc ultrices aliquam nunc. Vivamus facilisis pellentesque velit. Nulla nunc velit, vulputate dapibus, vulputate id, mattis ac, justo. Nam mattis elit dapibus purus. Quisque enim risus, congue non, elementum ut, mattis quis, sem. Quisque elit.

PUSTAKA

- [1] N. L. R. M. Rani, "Persepsi jurnalis dan praktisi humas terhadap nilai berita," 2013.
- [2] Wikipedia, "Berita bohong," diakses 27 November 2020. [Online]. Available: https://id.wikipedia.org/wiki/Berita_bohong
- [3] K. Kominfo, "Temuan isu hoaks," 03 2020. [Online]. Available: https://eppid.kominfo.go.id/storage/uploads/2_12_Data_Statistik_Hoax_Agustus_2018_-_31_Maret_2020.pdf
- [4] —, "Laporan isu hoax juni 2020," 07 2020. [Online]. Available: https://eppid.kominfo.go.id/storage/uploads/2_31_Laporan_Isu_Hoaks_Bulan_Juni_2020.pdf
- [5] M. L. Khan and I. Idris, "Recognize misinformation and verify before sharing: A reasoned action and information literacy perspective," *Behaviour and Information Technology*, 01 2019.
- [6] K. Wibowo, D. Rahmawan, and E. Maryani, "Penelitian di indonesia: umur tidak mempengaruhi kecenderungan orang menyebarkan hoaks," 2019, diakses 27 November 2020. [Online]. Available: <https://theconversation.com/penelitian-di-indonesia-umur-tidak-mempengaruhi-kecenderungan-orang-menyebarkan-hoaks-110621>

- [7] A. Aggarwal, A. Chauhan, D. Kumar, M. Mittal, and S. Verma, "Classification of fake news by fine-tuning deep bidirectional transformers based language model," p. 163973, 04 2020.
- [8] P. Bahad, P. Saxena, and R. Kamal, "Fake news detection using bi-directional lstm-recurrent neural network," *Procedia Computer Science*, vol. 165, pp. 74–82, 02 2020.
- [9] A. B. Prasetyo, R. R. Isnanto, D. Eridani, Y. A. A. Soetrisno, M. Arfan, and A. Sofwan, "Hoax detection system on indonesian news sites based on text classification using svm and sgd," in *2017 4th International Conference on Information Technology, Computer, and Electrical Engineering (ICITACEE)*, 10 2017.
- [10] F. Rahutomo, I. Pratiwi, and D. Ramadhani, "Eksperimen naïve bayes pada deteksi berita hoax berbahasa indonesia," *JURNAL PENELITIAN KOMUNIKASI DAN OPINI PUBLIK*, vol. 23, 07 2019.
- [11] "Cetak biru space x," 2021. [Online]. Available: <https://id.pinterest.com/pin/344032859037191430/>