Automatic Indonesian Hoax News Detection Using BERT

Reza Fuad Rachmadi
Department of Computer Engineering
Faculty of Intelligent Electrical
and Informatics Technology
Sepuluh Nopember
Institute of Technology
Surabaya, Indonesia 60111
fuad@te.its.ac.id

Mauridhi Hery Purnomo
Department of Computer Engineering
Faculty of Intelligent Electrical
and Informatics Technology
Sepuluh Nopember
Institute of Technology
Surabaya, Indonesia 60111
hery@ee.its.ac.id

Aufa Nabil Amiri
Department of Computer Engineering
Faculty of Intelligent Electrical
and Informatics Technology
Sepuluh Nopember
Institute of Technology
Surabaya, Indonesia 60111
aufa.17072@mhs.its.ac.id

***Abstrak*—Fake news or called hoax, is one of the things that still plaguing Indonesia. Even more so, with the rise of the social media, a fake news can spread wider and faster than ever before. Worse, Indonesian people have quite a high tendencies to share fake news. That is why, we are in a dire need of a method to detect fake news. This research is using BERT to automatically classifiy whether a news is a hoax or not. From a raw text, we applied a tokenization process before inputting the text to the BERT. Next, the pooled output of the BERT is being used as the input for Linear Regression, a tested-and-true method for classifying task. The output of the Linear Regression is then being used as a way to determine whether a news is a hoax or not. The purpose of this research is to create a machine learning model to help the people to determine whether a text is a fake news or not. The result of this research is a model to classify a hoax text with 89% in accuracy.**

***Kata kunci*—BERT, Hoax, Fake News Classification, Linear Regression**

I. Introduction

News is a report or a factual story, designed to be the fastest, has a good way of describing problems, and is just by nature to all problem in which it is choose to be published [1]. News also has a very important role in the public, not only because it is a good way to attain a new information, but also to broaden one knowledge.

Hoax or fake news is a way or method to try to deceive people so they believed something that is can't be considered correct and those incorrect things is more often than not is something only a mad-man would believe [2]. Not only reading a fake news will cost you your knowledge, hoax can have many other effects, ranging from the loss of reputation, money, up to even death threat.

Based on the data that we got from the Ministry of Communication and Informatics, there are a total of 5156 hoaxes that have been found only from a short range of August 2018 to March 2020. From January 2020 to March 2020, there are as many as 959 fake news that have been found [3]. Still based on the very same source, at Juny 2020, there are dozens of new hoaxes have been discovered every single day [4].

Nowadays, there are high chance everyone has a few social media account rather than those that are not. This in turn, has quite an effect on the spreading of the fake news, with those hoaxes spreading far and wide with speed never seen before. Based on the survey conducted by Khan and Idris, there are more than half of Indonesian people has a high tendency to share news links without feeling the need to do any validation of said news beforehand [5]. Another survey with similar topic conducted by Kunto with 480 respondence at East Java, shows that around 30% of the total of the responder has a tendency to share a fake news from mild to severe [6]. From those studies, it is safe to assume that Indonesian people in general, has a high tendency to share fake news through their social media accounts.

Neural networks is one of the many branches of machine learning study in which it is applying neurons, just like those that is usually found in human brain structure. Those neurons is used by neural network to process data which in turn resulting in an output. One of the newest things in neural network branch is a method called Bi-Directional Encoder Representations from Transformers or BERT in short. BERT is a method to get a context from a raw text in which it is inputted.

There are many previous works on this automatic hoax detection topic that have been done by other researchers in the past. Aggarway et al. has done an extensive research to see the difference between BERT, XGBoost and LSTM to classify fake news from english sources. Based on that research, turn out BERT has quite an edge to detect hoaxes compared to the other two method [7]. Another researcher under the name Bahad et al. has done another research to see which one is better between CNN, RNN, uni-directional LSTM RNN and bi-directional LSTM RNN when used also to classifying fake news. The result shows that LSTM coupled with attention-span, whether it is a uni-directional or bi-directional one, has quite a high accuracy compared to the other method like CNN or RNN [8]. From either of those two researchs, it can be concluded that if an algorithm is able to "remember" or know the context of the text, it will most likely has a higher accuracy if compared to the other non-"remember" approach.

But, if we are talking about Indonesia news detection state nowadays, there are not that many researcher has been doing that topic. There has been a research, done by Prasetijo et al. , that try to use SVM and SGD to detect Indonesian hoax news and resulting in a model with the accuracy of 85% [9]. Another research by Rahutomo et al. on the same topic but using naive bayes as the method, has been succesfully attain 80% accuracy on the same task [10].

The purpose of this research is to develop a model to automatically detect Indonesian hoax news by using BERT. The reason being is that by using BERT, hopefully, there will be an increase in efficiency and accuracy of Indonesian hoax

news detection.

II. Penelitian Terkait

Beberapa penelitian lain pernah dilakukan seperti yang dirumuskan oleh **?**] bahwa Fusce mauris. Vestibulum luctus nibh at lectus. Sed bibendum, nulla a faucibus semper, leo velit ultricies tellus, ac venenatis arcu wisi vel nisl. Vestibulum diam. Aliquam pellentesque, augue quis sagittis posuere, turpis lacus congue quam, in hendrerit risus eros eget felis. Maecenas eget erat in sapien mattis porttitor. Vestibulum porttitor. Nulla facilisi. Sed a turpis eu lacus commodo facilisis. Morbi fringilla, wisi in dignissim interdum, justo lectus sagittis dui, et vehicula libero dui cursus dui. Mauris tempor ligula sed lacus. Duis cursus enim ut augue. Cras ac magna. Cras nulla. Nulla egestas. Curabitur a leo. Quisque egestas wisi eget nunc. Nam feugiat lacus vel est. Curabitur consectetuer. Hasil tersebut kemudian menjadi persamaan 1.

$$\sum \mathbf{F} = 0 \;\Leftrightarrow\; \frac{\mathrm{d}\mathbf{v}}{\mathrm{d}t} = 0. \tag{1}$$

Suspendisse vel felis. Ut lorem lorem, interdum eu, tincidunt sit amet, laoreet vitae, arcu. Aenean faucibus pede eu ante. Praesent enim elit, rutrum at, molestie non, nonummy vel, nisl. Ut lectus eros, malesuada sit amet, fermentum eu, sodales cursus, magna. Donec eu purus. Quisque vehicula, urna sed ultricies auctor, pede lorem egestas dui, et convallis elit erat sed nulla. Donec luctus. Curabitur et nunc. Aliquam dolor odio, commodo pretium, ultricies non, pharetra in, velit. Integer arcu est, nonummy in, fermentum faucibus, egestas vel, odio.

Sed commodo posuere pede. Mauris ut est. Ut quis purus. Sed ac odio. Sed vehicula hendrerit sem. Duis non odio. Morbi ut dui. Sed accumsan risus eget odio. In hac habitasse platea dictumst. Pellentesque non elit. Fusce sed justo eu urna porta tincidunt. Mauris felis odio, sollicitudin sed, volutpat a, ornare ac, erat. Morbi quis dolor. Donec pellentesque, erat ac sagittis semper, nunc dui lobortis purus, quis congue purus metus ultricies tellus. Proin et quam. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos hymenaeos. Praesent sapien turpis, fermentum vel, eleifend faucibus, vehicula eu, lacus.

III. Arsitektur

A. Cetak Biru Roket

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, consectetuer at, consectetuer 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.

Morbi luctus, wisi viverra faucibus pretium, nibh est placerat odio, nec commodo wisi enim eget quam. Quisque libero justo, consectetuer 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

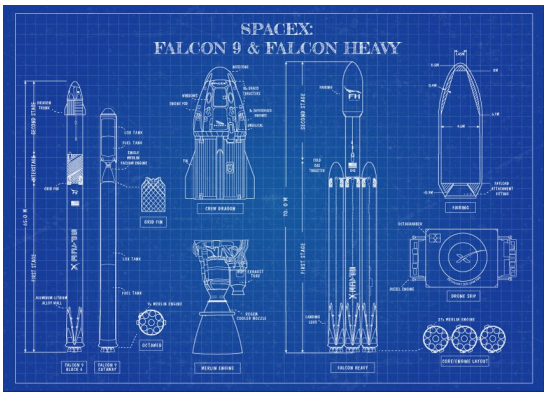


Fig. 1. Cetak biru roket yang akan diuji coba. [**?**]

TABLE I
Contoh tabel sederhana

Heading1	Heading2	Heading3
One	Two	Three
Four	Five	Six

orci luctus et ultrices posuere cubilia Curae; Pellentesque sit amet pede ac sem eleifend consectetuer. 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.

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 consectetuer odio sem sed wisi.

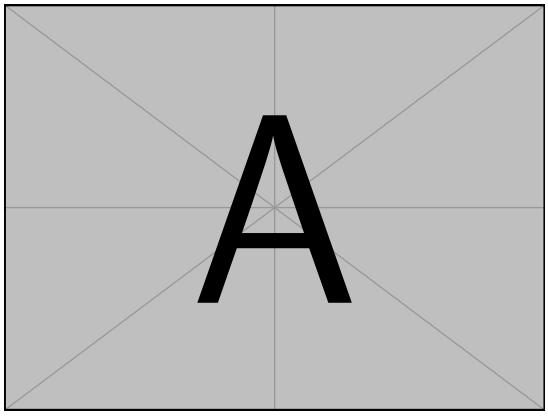
B. Lorem Ipsum

Sed feugiat. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Ut pellentesque augue sed urna. Vestibulum diam eros, fringilla et, consectetuer 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.

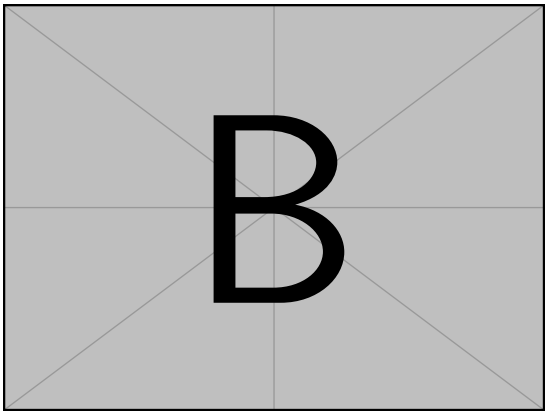
```
#include <iostream>

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 consectetuer tortor sapien facilisis magna. Mauris quis magna varius nulla scelerisque imperdiet.



(a) Hasil A



(b) Hasil B

Fig. 2. Contoh input beberapa gambar.

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. Vestibulum 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, consectetuer 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, consectetuer 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, consectetuer 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.

References

[1] Rani and N. L. R. Maha, "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. Prasetijo, 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.