IJCCS (Indonesian Journal of Computing and Cybernetics Systems)
Vol.15, No.3, July 2021, pp. 317~326
ISSN (print): 1978-1520, ISSN (online): 2460-7258
DOI: https://doi.org/10.22146/ijccs.66205

Transfer Learning of Pre-trained Transformers for Covid-19 Hoax Detection in Indonesian Language

Lya Hulliyyatus Suadaa*[1], Ibnu Santoso[2], Amanda Tabitha Bulan Panjaitan[3]
[1,2,3]Program Studi Komputasi Statistik, Politeknik Statistika STIS, Jakarta, Indonesia
e-mail: ***[1]lya@stis.ac.id**, [2]ibnu@stis.ac.id, [3]221709539@stis.ac.id

Abstrak

Pada saat ini, internet menjadi sumber berita paling populer. Akan tetapi, validitas dari artikel berita online sulit dinilai, apakah artikel tersebut berupa fakta atau hoaks. Hoaks terkait Covid-19 membawa efek problematik terhadap kehidupan manusia. Sistem pendeteksi hoaks yang akurat menjadi penting untuk menyaring informasi yang tersebar di internet. Pada penelitian ini, sistem pendeteksi hoaks diusulkan dengan menerapkan transfer learning dari pre-trained transformer model. Fine-tuned original pre-trained BERT, multilingual pre-trained mBERT, dan monolingual pre-trained IndoBERT digunakan untuk menyelesaikan tugas klasifikasi pada sistem pendeteksi hoaks. Berdasarkan hasil eksperimen, model fine-tuned IndoBERT yang dilatih di korpus monolingual berbahasa Indonesia lebih baik akurasinya dibandingkan fine-tuned original dan multilingual BERT dengan versi uncased. Akan tetapi, fine-tuned model mBERT dengan versi cased yang dilatih di corpus yang lebih besar mencapai kinerja yang paling baik dibandingkan dengan model lainnya.

Kata kunci*—deteksi hoaks, transfer learning, pre-trained transformer, pemrosesan teks berbahasa Indonesia*

Abstract

Nowadays, internet has become the most popular source of news. However, the validity of the online news articles is difficult to assess, whether it is a fact or a hoax. Hoaxes related to Covid-19 brought a problematic effect to human life. An accurate hoax detection system is important to filter abundant information on the internet. In this research, a Covid-19 hoax detection system was proposed by transfer learning of pre-trained transformer models. Fine-tuned original pre-trained BERT, multilingual pre-trained mBERT, and monolingual pre-trained IndoBERT were used to solve the classification task in the hoax detection system. Based on the experimental results, fine-tuned IndoBERT models trained on monolingual Indonesian corpus outperform fine-tuned original and multilingual BERT with uncased versions. However, the fine-tuned mBERT cased model trained on a larger corpus achieved the best performance.

Keywords—*hoax detection, transfer learning, pre-trained transformer, Indonesian language text processing*

Received May 30th,2021; Revised July 31th, 2021; July 31th, 2021

1. INTRODUCTION

Nowadays, hoaxes can spread easily through the internet, as one of the dark sides of information technology development. Information technology facilitates people to connect with others and exchange information without constraint. Information that can be consumed is also increasing in terms of quantity and variety. However, the information that is exchanged is often false, inaccurate, and maybe deliberately distributed with a specific purpose, called a hoax.

There are several definitions of a hoax that are not much different from various dictionaries [13] [14] [15]. Principally, a hoax is false information, regardless of the purpose of disseminating the information. In Indonesia, it is indicated that hoax information comes from 800 thousand websites [16].

Hoaxes related to covid-19 have been circulating in the community through social media and various news sites. More than 2,300 reports of hoaxes and conception theories about covid-19 have been recorded. Misinformation about covid-19 has affected the lives of at least 800 people worldwide [17]. Considering the problematic effects of covid-19 hoaxes, we intend to solve hoax detection problems by classifying articles into a hoax and fact.

Several studies have elaborated hoax detection tasks and proposed using classic classification models such as K Nearest Neighbor (KNN) [10], naive Bayes classifier [11][18], and random forest [1]. These models need manual feature engineering by defining representative features beforehand. However, a deeper analysis is needed to decide on better features. Overcoming the feature engineering problem, a deep learning model was introduced to solve classification tasks without feature engineering. Features were extracted through their word embeddings representing the meaning of each token in the input texts. A recent study has been proved that deep learning models are superior to classic classifiers in a hoax detection system [12].

Due to the complexity of deep learning architecture, deep learning models require a lot of training data. However, constructing a large dataset, especially for supervised tasks, is expensive. Building deep learning models using such dataset also needs more considerable resources, such as high-performance computer architecture. Avoiding the need to train a new model from scratch, public pre-trained language models were proposed.

Recent studies have shown that pre-trained models trained on a large corpus can be successfully solved various downstream natural language processing tasks by transfer learning. One of the pre-trained language representation models outperforming many task-specific architectures is BERT. BERT, Bidirectional Encoder Representation from Transformers, is designed to pre-train deep representation of texts from both left and right directions [2]. BERT can be easily fine-tuned for a classification task by simply adding one additional classification layer, thus avoiding the need to train a new model from scratch. Since our dataset is limited, we aim to take advantage of a larger pre-trained language representation by transfer learning the pre-trained models to our article classification task and develop a accurate hoax detection system.

The promising results of BERT architecture in obtaining deeper context representation of input texts encourage the development of BERT trained on the different corpus. Devlin et al. also offered a multilingual BERT (mBERT) [2], a pre-trained BERT trained on Wikipedia documents for 104 languages, achieving impressive performance for zero-shot cross-lingual transfer [3]. Several studies incorporated mBERT in Indonesian language processing tasks, such as aspect-based sentiment analysis in Indonesian review datasets [7]. Even the original pre-trained BERT was elaborated in other Indonesian tasks, such as hate speech classification [8] and summarization [9].

Recently, two different versions of BERT trained on Indonesian corpus have been released in the same year, called IndoBERT. The first version of IndoBERT was proposed by the IndoNLU team, which is trained on a large and clean Indonesian dataset (Indo4B) collected from publicly available sources such as social media texts, blogs, news, and websites [4].

Another version of IndoBERT, proposed by the IndoLEM team, is trained on Indonesian Wikipedia, news articles (Kompas, Tempo, and Liputan6), and Web Corpus [5]. These monolingual BERT trained on the Indonesian language corpus encourage further research exploring transfer learning of pre-trained BERT models to various Indonesian language processing tasks.

In this study, we elaborated the original pre-trained BERT, the multilingual BERT (mBERT) and the monolingual IndoBERT to develop a hoax detection system and presented the experimental results. The performance of each model was compared and analyzed for selecting the hoax detection system with the best accuracies.

2. METHODS

We proposed transfer learning by finetuning pre-trained transformers models for hoax detection tasks. A flow chart of our proposed system is illustrated in Figure 1.

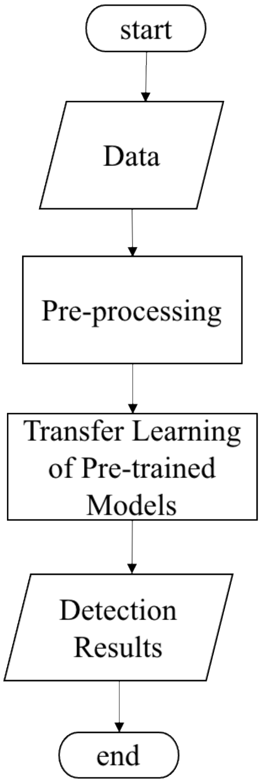
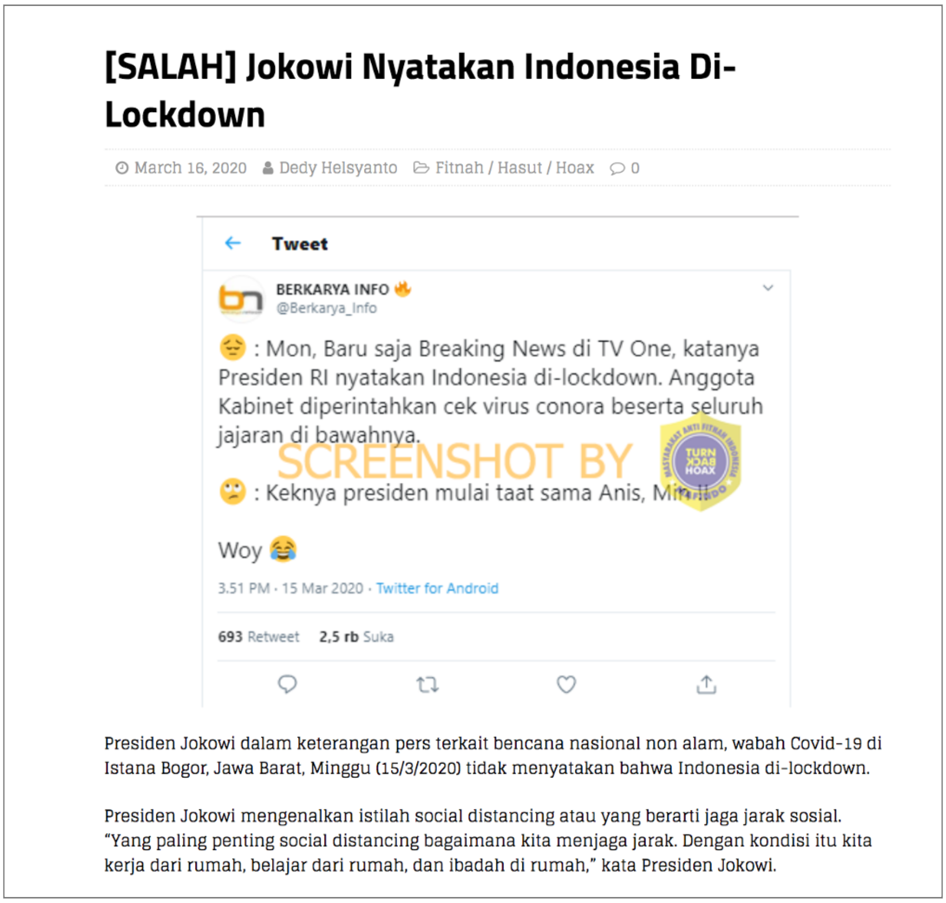


Figure 1 Hoax Detection Phases with Transfer Learning

2.1 Dataset

We used a dataset of Covid-19 articles in Indonesian languages collected by [1] consisted of hoax articles from Turnbackhoax.id and fact articles from Detik.com. Keywords that were used for selecting the Covid-19 articles in this study are “covid”, “corona”, and “pandemik”. Page interfaces of hoax and fact article examples are shown in Figure 2 and Figure 3, respectively.



[SALAH] Jokowi Nyatakan Indonesia Di-Lockdown

March 16, 2020 Dedy Helsyanto Fitnah / Hasut / Hoax 0

Tweet

BERKARYA INFO 🔥
@Berkarya_Info

😔 : Mon, Baru saja Breaking News di TV One, katanya Presiden RI nyatakan Indonesia di-lockdown. Anggota Kabinet diperintahkan cek virus conora beserta seluruh jajaran di bawahnya.

🙄 : Keknya presiden mulai taat sama Anis, Min!!

Woy 😂

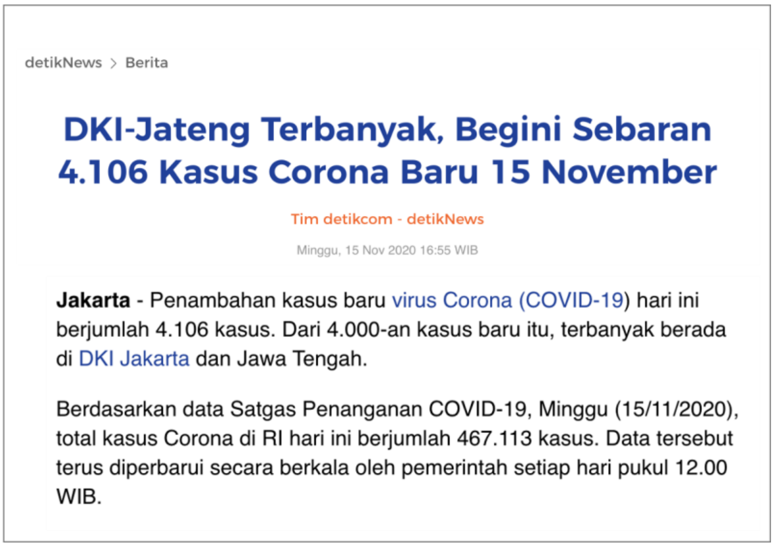
3.51 PM · 15 Mar 2020 · Twitter for Android

693 Retweet 2,5 rb Suka

Presiden Jokowi dalam keterangan pers terkait bencana nasional non alam, wabah Covid-19 di Istana Bogor, Jawa Barat, Minggu (15/3/2020) tidak menyatakan bahwa Indonesia di-lockdown.

Presiden Jokowi mengenalkan istilah social distancing atau yang berarti jaga jarak sosial. "Yang paling penting social distancing bagaimana kita menjaga jarak. Dengan kondisi itu kita kerja dari rumah, belajar dari rumah, dan ibadah di rumah," kata Presiden Jokowi.

Figure 2 A Page Example of Hoax Article



detikNews > Berita

DKI-Jateng Terbanyak, Begini Sebaran 4.106 Kasus Corona Baru 15 November

Tim detikcom - detikNews

Minggu, 15 Nov 2020 16:55 WIB

Jakarta - Penambahan kasus baru virus Corona (COVID-19) hari ini berjumlah 4.106 kasus. Dari 4.000-an kasus baru itu, terbanyak berada di DKI Jakarta dan Jawa Tengah.

Berdasarkan data Satgas Penanganan COVID-19, Minggu (15/11/2020), total kasus Corona di RI hari ini berjumlah 467.113 kasus. Data tersebut terus diperbarui secara berkala oleh pemerintah setiap hari pukul 12.00 WIB.

Figure 3 A Page Example of Fact Article

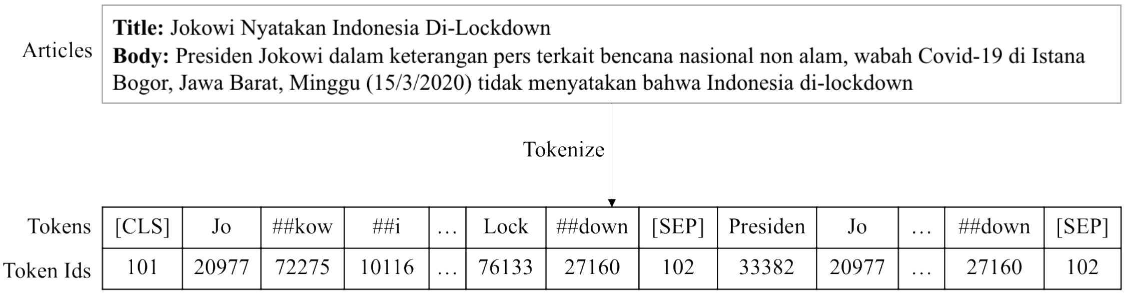
A classification model was developed to detect whether an article is a hoax or fact. Examples of articles and their classes in the dataset are shown in Table 1.

Table 1 Examples of articles and their classes in dataset

No	Article		Class
	Title	Body	
1	Jokowi Nyatakan Indonesia Di-Lockdown	Presiden Jokowi dalam keterangan pers terkait bencana nasional non alam, wabah Covid-19 di Istana Bogor, Jawa Barat, Minggu (15/3/2020) tidak menyatakan bahwa Indonesia di-lockdown…	Hoax
2	Ternyata virus corona dapat diobati dengan cara berendam di AIR LAUT	Tidak bisa disembuhkan hanya dengan berendam di laut. Pasalnya, virus Corona tidak menyerang permukaan tubuh seperti kulit, melainkan menyerang sel-sel di dalam tubuh…	Hoax
3	DKI-Jateng Terbanyak, Begini Sebaran 4.106 Kasus Corona Baru 15 November	Jakarta - Penambahan kasus baru virus Corona (COVID-19) hari ini berjumlah 4.106 kasus. Dari 4.000-an kasus baru itu, terbanyak berada di DKI Jakarta dan Jawa Tengah…	Fact
4	Pemprov Kaltim Dapat Jatah 2,2 Juta Vaksin COVID-19	Samarinda - Provinsi Kalimantan Timur (Kaltim) mendapatkan bantuan sebanyak 2,2 juta vaksin COVID-19 dari pemerintah pusat. Provinsi Kaltim sebelumnya tidak masuk daftar 10 daerah yang menjadi prioritas penerima bantuan vaksin virus Corona…	Fact

2. 2 Pre-processing

We used a pre-trained tokenizer to transform texts into sub-word tokens avoiding out-of-vocabulary problems. Following [2], [CLS] token is added at the beginning of articles tokens and [SEP] at the end of tokens. Then, we separated the title and body of articles into two segments by inserting a [SEP] token. Then, all tokens were transformed into token ids. An example of our tokenization process in pre-processing phase is illustrated in Figure 4.



Articles

Title: Jokowi Nyatakan Indonesia Di-Lockdown
Body: Presiden Jokowi dalam keterangan pers terkait bencana nasional non alam, wabah Covid-19 di Istana Bogor, Jawa Barat, Minggu (15/3/2020) tidak menyatakan bahwa Indonesia di-lockdown

Tokens	[CLS]	Jo	##kow	##i	…	Lock	##down	[SEP]	Presiden	Jo	…	##down	[SEP]
Token Ids	101	20977	72275	10116	…	76133	27160	102	33382	20977	…	27160	102

Figure 4 An Example of Tokenization Process

In our study, we fed the original texts as our inputs to the tokenizer. However, when uncased models were used, we did case-folding by reducing all letters to lowercase.

2. 3 Transfer Learning of Pre-trained Models

We fine-tuned a pre-trained transformers BERT to obtain context representation of our input texts. We adapt fine-tuned BERT architecture in [2] to solve the classification task for our hoax detection system. Our proposed fine-tuned architecture is depicted in Figure 5.

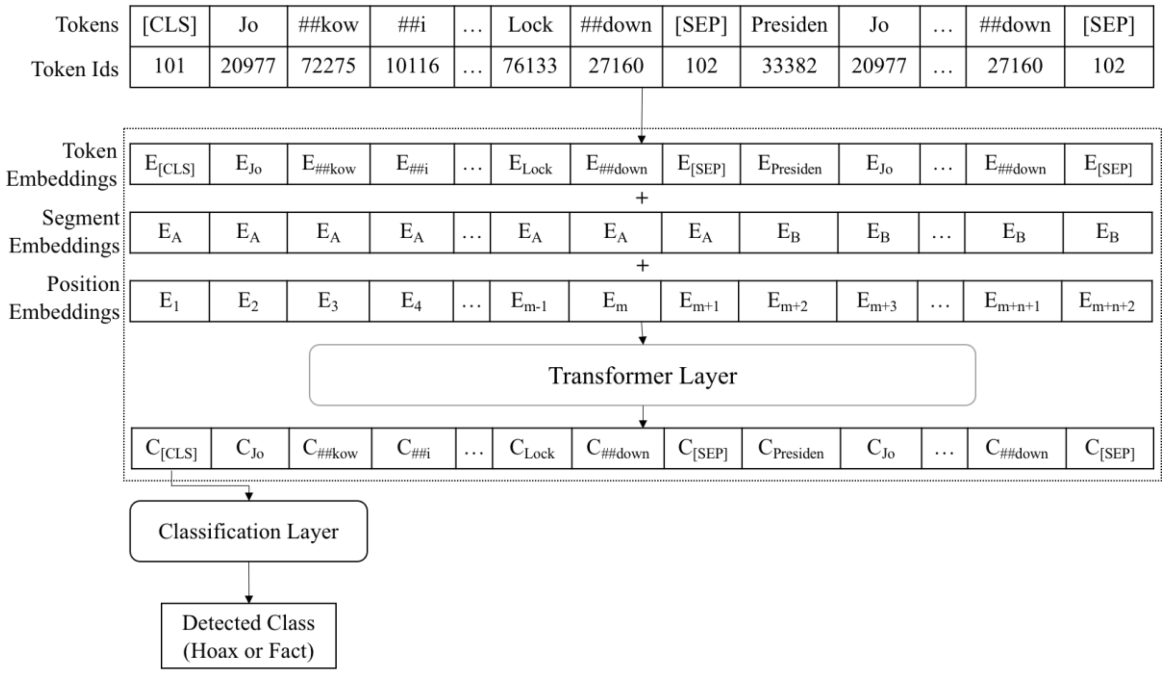


Figure 5 Fine-tuned BERT Architecture for Hoax Detection

We assigned token embeddings representing the meaning of each token, segment embeddings to discriminate the title and body of the article, and position embeddings covering the token position in our input sequences. The summation of these embeddings was fed to the Transformer layer of BERT. We used the top context [CLS] token as a representation of sequence tokens. Then, we added a classification layer to detect whether an article is a hoax or not. We used several BERT models trained on different corpus: original BERT, multilingual BERT, and IndoBERT.

2. 3.1 BERT

An original BERT model is trained on the BooksCorpus (800M words) and Wikipedia (2,500M words) [2]. We used BERT-based-cased and BERT-based-uncased in our experiments.

2. 3.2 Multilingual BERT

A multilingual BERT (mBERT) is trained on Wikipedia documents for 104 languages, including Indonesian and has been efficiently fine-tuned for document classification in several languages [2][3]. mBERT-base-cased and mBERT-base-uncased were used in our study.

2. 3.3 IndoBERT

There are two kinds of IndoBERT trained on a different corpus, proposed by IndoNLU [4] and IndoLEM [5] teams. IndoBERT of the IndoNLU is trained on around four billion words of Indonesian pre-processed text data (≈ 23 GB) from publicly available sources such as social media texts, blogs, news, and websites [4]. IndoBERT of the IndoLEM is trained on Indonesian Wikipedia (74M words), Indonesian news articles (55M words), and an Indonesian Web Corpus (90M words) [5]. Both models provided only the uncased models.

3. RESULTS AND DISCUSSION

We implemented our models in Pytorch and used a transformer library built by the Huggingface team [6]. For optimization in the fine-tuned phase, we used Adam as the optimizer with a batch size of 32 and a learning rate of 3×10^{-6}. As evaluation metrics, we reported accuracy, precision, recall, and F1 scores.

We fine-tuned the models for seven epochs. Based on our experimental results, the models tended to overfit after the seventh epoch. The improvement of accuracy scores in the fine-tuned phase of our proposed models is shown in Figure 6. We validate our models for each epoch in the train and test set. No validation set is available in this dataset.

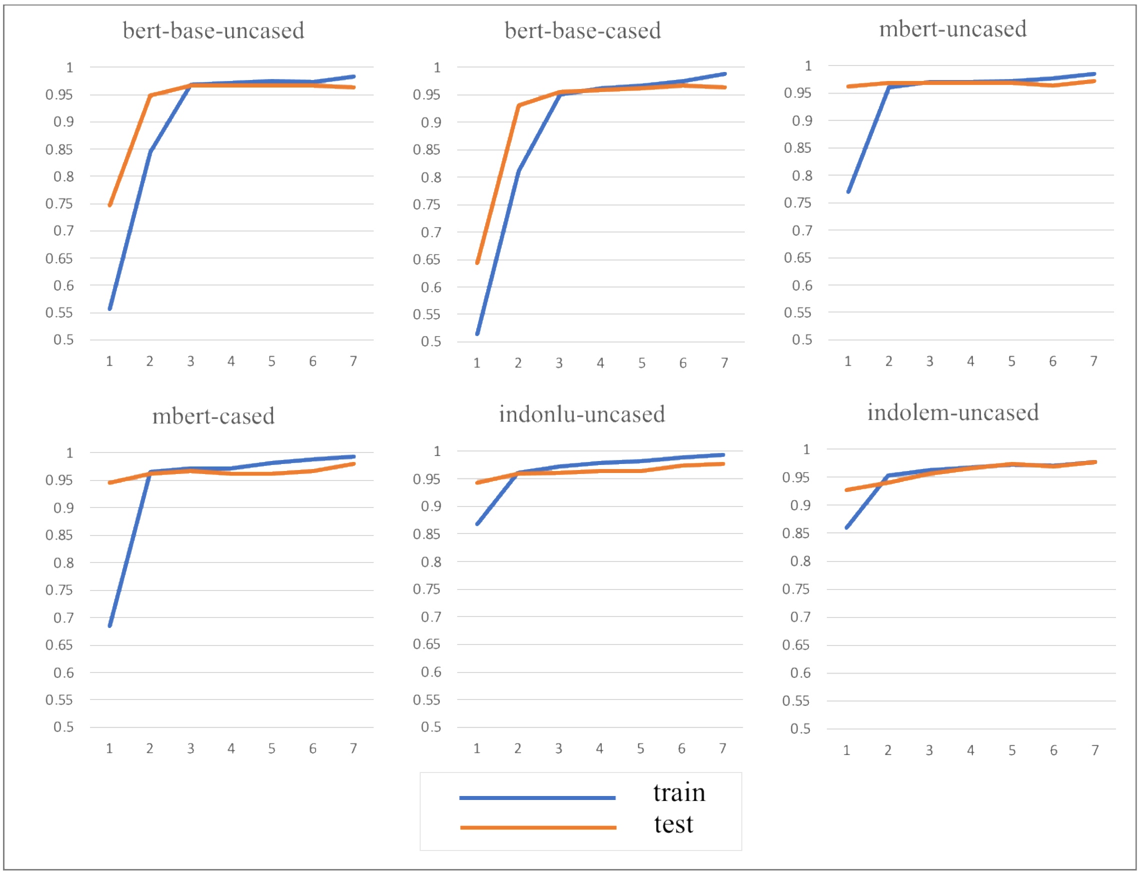


Figure 6 Accurracies of Our Proposed Models for Each Epoch in Train and Test Sets

As reported in Figure 6, the accuracies of all models were improved after each epoch. It proves that transfer learning from the pre-trained models increases the model performances. We successfully took advantage of pre-trained models trained on a large corpus by fine-tuning the models in our task, even with a limited dataset. The accuracy scores of fine-tuned BERT in the testing sets were sharply boosted in the first three epochs, then continue with a slight increase for the next epoch. Unlike the original fine-tuned BERT models, fine-tuned mBERT and IndoBERT models were slightly increased from the first epoch, but the accuracy score was started from more than 90%.

We compared our proposed models with transfer learning to Random Forest without feature engineering and with feature engineering, the best classic classification models for this task reported in [1]. Our proposed models did not need feature engineering since text representations were automatically obtained through their token embeddings. Our experimental results are shown in Table 2.

Table 2 Experimental results

Models	Accuracy	Precision	Recall	F1
Random Forest w/o Feature Engineering [1]	93.79	89.25	98.81	93.79
Random Forest w Feature Engineering [1]	96.05	92.31	**100**	96
Fine-tuned BERT-base-uncased	96.38	96.57	96.37	96.38
Fine-tuned BERT-base-cased	96.38	96.38	96.47	96.38
Fine-tuned mBERT-base-uncased	97.16	97.27	97.15	97.16
Fine-tuned mBERT-base-cased	**97.93**	**97.93**	97.96	**97.93**
Fine-tuned IndoBERT-base-uncased (IndoNLU)	97.67	97.74	97.67	97.67
Fine-tuned IndoBERT-base-uncased (IndoLEM)	97.67	97.78	97.67	97.67

As seen in Table 2, our fine-tuned models achieved better accuracy, precision, and F1 scores than reported in the previous works. The original fine-tuned BERT with uncased and cased models gave a similar performance with 96.38 of accuracy and F1 scores. Unlike the original BERT, the fine-tuned mBERT with cased model reported better performance than the uncased one. Since the mBERT model was trained on a larger corpus in various languages, differentiate capital and not capital letters of article texts improved our hoax detection performances. The cased version models were efficiently handling words with capital letters by using different embeddings.

Both fine-tuned IndoBERT models, IndoNLU and IndoLEM, provided only an uncased version. Both models achieved similar performance with 97.67 accuracies and outperformed fine-tuned BERT and mBERT uncased models. IndoBERT as a monolingual pre-trained model gave a better score than mBERT, a multilingual one. However, the cased version of fine-tuned mBERT model outperformed all fine-tuned models since mBERT-cased trained on a larger corpus than others.

4. CONCLUSIONS

We proposed transfer learning of pre-trained models to solve the COVID-19 hoax detection task. We fine-tuned original pre-trained BERT, multilingual pre-trained mBERT, and monolingual pre-trained IndoBERT to our classification task and reported the results. Our fine-tuned IndoBERT models trained on monolingual Indonesian corpus outperformed fine-tuned original and multilingual BERT with uncased versions. However, the fine-tuned mBERT cased model trained a larger corpus achieved the best performance.

REFERENCES

[1] A. T. B. Panjaitan and I. Santoso, “Deteksi Hoaks Pada Berita Berbahasa Indonesia Seputar COVID-19,” *Jurnal FORMAT (Teknik Informatika).*, vol. 10, no. 1, p. 76, 2021 [Online]. Available: https://publikasi.mercubuana.ac.id/index.php/format/article/view/10978. [Accessed: 26-May-2021]

[2] J. Devlin, M. Chang, K. Lee, and K. Toutanova, “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding,” *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, p. 4171–4186, 2019 [Online]. Available: http https://www.aclweb.org/anthology/N19-1423/. [Accessed: 26-May-2021]

[3] S. Wu and M. Dredze, “Beto, Bentz, Becas: The surprising cross-lingual effectiveness of BERT,” *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing*, pp. 833-844, 2019 [Online]. Available: https://www.aclweb.org/anthology/D19-1077/. [Accessed: 26-May-2021]

[4] B. Wilie, K. Vincentio, G. I. Winata, S. Cahyawijaya, X. Li, Z. Y. Lim, S. Soleman, R. Mahendra, P. Fung, S. Bahar, and A. Purwarianti, “IndoNLU: Benchmark and Resources for Evaluating Indonesian Natural Language Understanding,” *Proceedings of the 1st Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 10th International Joint Conference on Natural Language Processing*, pp. 843-857, 2020 [Online]. Available: https://www.aclweb.org/anthology/2020.aacl-main.85/. [Accessed: 26-May-2021]

[5] F. Koto, A. Rahimi, J. H. Lau, and T. Baldwin, “IndoLEM and IndoBERT: A Benchmark Dataset and Pre-trained Language Model for Indonesian NLP,” *Proceedings of the 28th International Conference on Computational Linguistics*, pp. 757-770, 2020 [Online]. Available: https://www.aclweb.org/anthology/2020.coling-main.66/. [Accessed: 26-May-2021]

[6] T. Wolf, L. Debut, V. Sanh, J. Chaumond, C. Delangue, A. Moi, P. Cistac, T. Rault, R. Louf, M. Funtowicz, J. Davison, S. Shleifer, P. Platen, C. Ma, Y. Jernite, J. Plu, C. Xu, T. L. Scao, S. Gugger, M. Drame, Q. Lhoest, and A. Rush. “Transformers: State-of-the-Art Natural Language Processing,” *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pp. 38-45 [Online]. Available: https://www.aclweb.org/anthology/2020.emnlp-demos.6/ [Accessed: 26-May-2021]

[7] A. N. Azhar and M. L. Khodra, “Fine-tuning Pretrained Multilingual BERT Model for Indonesian Aspect-based Sentiment Analysis,” *Proceedings of the 7th International Conference on Advance Informatics: Concepts, Theory and Applications (ICAICTA 2020)*, 2020 [Online]. Available: https://ieeexplore.ieee.org/document/9428882. [Accessed: 26-May-2021]

[8] Ilham Firdausi Putra; Ayu Purwarianti, “Improving Indonesian Text Classification Using Multilingual Language Model,” *Proceedings of the 7th International Conference on Advance Informatics: Concepts, Theory and Applications (ICAICTA 2020)*, 2020 [Online]. Available: https://ieeexplore.ieee.org/document/9429038. [Accessed: 26-May-2021]

[9] R. Wijayanti; M. L. Khodra, and D. H. Widyantoro, “Indonesian Abstractive Summarization using Pre-trained Model,” *Proceedings of the 3rd East Indonesia Conference on Computer and Information Technology (EIConCIT 2021)*, 2021 [Online]. Available: https://ieeexplore.ieee.org/document/9431880. [Accessed: 26-May-2021]

[10] E. Zuliarso, M. T. Anwar, K. Hadiono and I. Chasanah, “Detecting Hoaxes in Indonesian News Using TF/TDM and K Nearest Neighbor,” *IOP Conference Series: Materials Science and Engineering,* Vol. 835, 2019 [Online]. Available:, https://iopscience.iop.org/article/10.1088/1757-899X/835/1/012036. [Accessed: 26-May-2021]

[11] I. Y. R. Pratiwi, R. A. Asmara, and F. Rahutomo, “Study of Hoax News Detection using Naïve Bayes Classifier in Indonesian Language,” *Proceedings of the 11th International Conference on Information & Communication Technology and System (ICTS)*, 2017 [Online]. Available: https://ieeexplore.ieee.org/document/8265649. [Accessed: 26-May-2021]

[12] B. P. Nayoga, R. Adipradana, R. Suryadia, and D. Suhartono, “Hoax Analyzer for Indonesian News Using Deep Learning Models,” *Procedia Computer Science: Special Issues of the 5th International Conference on Computer Science and Computational Intelligence*, 2020 [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1877050921000739. [Accessed: 26-May-2021]

[13] Cambridge Dictionary, “Definition of Hoax”, 2021 [Online]. Available: https://dictionary.cambridge.org/dictionary/english/hoax. [Accessed: 26-May-2021]

[14] Collins Dictionary, “Definition of Hoax”, 2021 [Online]. https://www.collinsdictionary.com/dictionary/english/hoax. [Accessed: 26-May-2021]

[15] Merriam Webster, “Definition of Hoax”, 2021 [Online]. https://www.merriam-webster.com/dictionary/hoax. [Accessed: 26-May-2021]

[16] Kominfo, “There are 800,000 Hoax Spreader Sites in Indonesia,” 12-Dec-2017 [Online], https://kominfo.go.id/content/detail/12008/ada-800000-situs-penyebar-hoax-di-indonesia/0/highlight_media. [Accessed: 26-May-2021]

[17] Forbes, “Report: More Than 800 Deaths And 5,800 Hospitalizations Globally May Have Resulted From COVID-19 Misinformation,” 23-August-2020 [Online]. https://www.forbes.com/sites/markhall/2020/08/23/coronavirus-misinformation/. [Accessed: 26-May-2021]

[18] M. Granik and V. Mesyura, "Fake news detection using naive Bayes classifier," *Proceedings of the 2017 IEEE First Ukraine Conference on Electrical and Computer Engineering (UKRCON),* pp. 900-903, 2017 [Online]. Available: https://ieeexplore.ieee.org/document/8100379. [Accessed: 26-May-2021]