

**SISTEM PENDETEKSI BERITA PALSU PADA KONTEN BERITA
BERBAHASA INDONESIA MENGGUNAKAN METODE *LONG SHORT-
TERM MEMORY (LSTM)***

SKRIPSI

Diajukan Untuk Memenuhi Sebagian dari
Syarat Memperoleh Gelar Sarjana Komputer
Program Studi Ilmu Komputer



oleh:

M. Rifky Maulana R

1904051

**PROGRAM STUDI ILMU KOMPUTER
DEPARTEMEN PENDIDIKAN ILMU KOMPUTER
FAKULTAS PENDIDIKAN MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS PENDIDIKAN INDONESIA
2023**

**SISTEM Pendeteksi Berita Palsu pada Konten Berita
Berbahasa Indonesia Menggunakan Metode *LONG SHORT-TERM
MEMORY (LSTM)***

Oleh

M. Rifky Maulana R

NIM 1904051

Sebuah Skripsi yang Diajukan untuk Memenuhi Salah Satu Syarat Memperoleh Gelar Sarjana Komputer pada Fakultas Pendidikan Matematika dan Ilmu Pengetahuan Alam

© M. Rifky Maulana R

Universitas Pendidikan Indonesia

Bulan 2023

Hak Cipta Dilindungi Undang-Undang

Skripsi ini tidak boleh diperbanyak seluruhnya atau sebagian, dengan dicetak ulang, difoto kopi, atau cara lainnya tanpa izin dari penulis

M RIFKY MAULANA R

SISTEM PENDETEKSI BERITA PALSU PADA KONTEN BERITA
BERBAHASA INDONESIA MENGGUNAKAN METODE *LONG SHORT-
TERM MEMORY (LSTM)*

Disetujui dan disahkan oleh pembimbing:

Pembimbing I,



Dr. Yudi Wibisono, S.T., M.T.

NIP. 197507072003121003

Pembimbing II,



Herbert Siregar, M.T.

NIP. 197005022008121001

Mengetahui,

Ketua Program Studi Ilmu Komputer

A handwritten signature in blue ink, which appears to be "Dr. Muhamad Nursalman".

Dr. Muhamad Nursalman, M.T.

NIP. 197909292006041002.

**SISTEM PENDETEKSI BERITA PALSU PADA KONTEN BERITA
BERBAHASA INDONESIA MENGGUNAKAN METODE *LONG SHORT-
TERM MEMORY (LSTM)***

ABSTRAK

Semakin berkembangnya teknologi membuat masyarakat dapat mengakses berita dimanapun dan kapanpun melalui internet maupun media sosial. Namun, berita yang tersebar di internet seringkali mengandung fakta yang direkayasa untuk menyesatkan pembacanya yang disebut sebagai berita palsu (hoaks). Berita palsu dapat menyebar dengan cepat dikarenakan ketidaktahuan masyarakat dalam menyaring informasi yang didapat dan tidak sadar menyebarkan berita palsu kepada orang lain. Oleh karena itu, perlu dibuat sebuah sistem yang dapat membantu masyarakat untuk membedakan antara berita palsu dan berita fakta. Pendekripsi berita palsu telah banyak dilakukan dengan menggunakan berbagai metode *deep learning* murni tanpa bantuan model *word embedding*. Namun, penelitian ini membuat sistem pendekripsi berita palsu dengan menggunakan algoritma *Long Short Term Memory (LSTM)* serta menerapkan model Indobert sebagai model *word embedding* untuk data input yang digunakan. Dalam penerapannya, penelitian ini menggunakan dataset yang berisi 3309 data teks narasi berita hoaks dan non-hoaks berbahasa Indonesia yang dikumpulkan dari situs turnbackhoax.id dan beberapa portal berita terpercaya serta website kementerian Indonesia. Penelitian ini juga membuat antarmuka sederhana yang dapat diakses oleh masyarakat untuk melakukan klasifikasi berita. Dari hasil pengujian yang telah dilakukan, model ini dapat memprediksi berita palsu dengan tingkat akurasi 94.2%, *precision* 0.943, *recall* 0.940, dan *f-1 score* 0.941 yang lebih baik daripada metode pembanding. Hasil penelitian ini dapat dikembangkan lagi untuk mendekripsi berita palsu pada berbagai *platform* media sosial.

Kata Kunci: Hoaks, LSTM, *hoax prediction system*, *deep learning*, *Indobert*.

HOAX DETECTION SYSTEM ON INDONESIAN NEWS CONTENT USING LONG SHORT-TERM MEMORY (LSTM) METHOD

ABSTRACT

The technology development allows people to access news anywhere and anytime through the internet and social media. However, news spread on the internet often contains facts that are engineered to mislead readers, which is called fake news (hoax). Fake news can spread quickly due to people's ignorance in filtering the information obtained and unconsciously spreading fake news to others. Therefore, it is necessary to create a system that can help people to distinguish between fake news and factual news. Fake news detection has been done using various pure deep learning methods without the help of word embedding models. However, this research makes a fake news detection system using the Long Short Term Memory (LSTM) algorithm and applies the Indobert model as a word embedding model for the input data used. In its application, this research uses a dataset containing 3309 hoax and non-hoax news narrative text data in Indonesian language collected from the turnbackhoax.id website and several trusted news portals and Indonesian ministry websites. This research also creates a simple interface that can be accessed by the public to classify news. From the test results, this model can predict fake news with an accuracy rate of 94.2%, precision 0.943, recall 0.940, and f-1 score 0.941 which is better than the comparison method. The results of this study can be further developed to detect fake news on various social media platforms.

Keywords: hoax, LSTM, hoax prediction system, deep learning, Indobert.

DAFTAR ISI

KATA PEGANTAR	ii
UCAPAN TERIMA KASIH.....	iii
ABSTRAK	iv
<i>ABSTRACT</i>	v
DAFTAR ISI.....	vi
DAFTAR GAMBAR	viii
DAFTAR TABEL.....	xi
1 BAB I PENDAHULUAN	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	6
1.3 Tujuan Penelitian.....	6
1.4 Manfaat Penelitian.....	6
1.5 Batasan Masalah.....	7
BAB II KAJIAN PUSTAKA	8
2.1 Berita Palsu.....	8
2.1.1 Definisi berita palsu	8
2.1.2 Faktor Penyebab Tersebarnya Berita Palsu	12
2.2 <i>Website Turnbackhoax.id</i>	13
2.3 <i>Text Pre-processing</i>	15
2.4 Word Embedding.....	17
2.5 Algoritma <i>Recurrent Neural Network</i> (RNN)	23
2.6 Algoritma Long Short-Term Memory (LSTM).....	24
2.6.1 Definisi LSTM	24
2.6.2 Pseudocode LSTM.....	31
2.7 Pengukuran Kinerja Model.....	38
2.8 Penelitian Sebelumnya	40
BAB III METODOLOGI PENELITIAN.....	46
3.1 Desain Penelitian	46
3.2 Lingkungan Implementasi	48
3.3 Teknik Analisis.....	49
BAB IV HASIL DAN PEMBAHASAN	50

4.1	Dataset	50
4.2	Text Pre-processing	52
4.3	Word Embedding.....	55
4.4	Pembuatan Sistem Pendekripsi Berita Palsu.....	57
4.4.1	Pembuatan Model LSTM.....	57
4.4.2	Pembuatan Model RNN.....	69
4.4.3	Antarmuka Sistem.....	72
4.5	Hasil dan Pembahasan	74
4.6	Permasalahan-Permasalahan dalam Pembuatan Sistem	77
BAB V KESIMPULAN, IMPLIKASI DAN SARAN		80
5.1	Kesimpulan.....	80
5.2	Implikasi	80
5.3	Saran	81
DAFTAR PUSTAKA		82
LAMPIRAN		87

DAFTAR GAMBAR

Gambar 1.1 Contoh berita palsu (kominfo.go.id)	1
Gambar 1.2 Ragam sumber informasi yang paling sering diakses masyarakat (KIC)	2
Gambar 1.3 Kebiasaan masyarakat dalam mengonsumsi berita online (KIC)	3
Gambar 2.1 Ilustrasi berita "The Baloon"	9
Gambar 2.2 Ilustrasi berita "Broadcasting the Barricades"	10
Gambar 2.3 Berita palsu informasi bencana alam di Indonesia (turnbackhoax.id)	11
Gambar 2.4 Halaman utama situs Turnbackhoax.id	13
Gambar 2.5 Halaman Artikel Berita Hoaks (Turnbackhoax.id)	14
Gambar 2.6 Tampilan isi salah satu halaman berita (Turnbackhoax.id)	14
Gambar 2.7 Tahapan Text Preprocessing	15
Gambar 2.8 Ilustrasi Struktur CBOW (Rong, 2014)	18
Gambar 2.9 Ilustrasi Struktur Skipgram (Rong, 2014)	19
Gambar 2.10 Arsitektur BERT (Devlin dkk., 2019)	21
Gambar 2.11 Framework Pretraining dan Fine tuning BERT (Devlin dkk., 2019)	22
Gambar 2.12 Arsitektur RNN	24
Gambar 2.13 Ilustrasi struktur gates (Olah, 2015)	25
Gambar 2.14 Modul perulangan RNN Standard dengan satu layer (Olah, 2015)	25
Gambar 2.15 Modul perulangan LSTM dengan empat layers (Olah, 2015)	25
Gambar 2.16 Ilustrasi cell states	26
Gambar 2.17 Ilustrasi langkah pertama LSTM (Olah, 2015)	26
Gambar 2.18 Ilustrasi Tahapan Kedua LSTM (Olah, 2015)	28
Gambar 2.19 Ilustrasi Tahap Ketiga LSTM (Olah, 2015)	29
Gambar 2.20 Ilustrasi Tahap Keempat LSTM (Olah, 2015)	29
Gambar 3.1 Diagram Alir Tahapan Penelitian	46
Gambar 4.1 Dataset yang digunakan	50
Gambar 4.2 Statistik data berdasarkan kategori berita	51
Gambar 4.3 Code Python data cleaning	52
Gambar 4.4 Code Python untuk filtering stopword	53

Gambar 4.5 Code Python untuk stemming	54
Gambar 4.6 Data hasil preprocessing	54
Gambar 4.7 Pembagian dataset	55
Gambar 4.8 Pretrained model Indobert	55
Gambar 4.9 Proses encoding data	56
Gambar 4.10 Contoh hasil encoding salah satu data training	57
Gambar 4.11 Code untuk mengambil data hasil preprocessing	57
Gambar 4.12 Code untuk mengambil data hasil preprocessing (tanpa stemming)	57
Gambar 4.13 Code untuk inisialisasi bobot dan bias	58
Gambar 4.14 Coding forward propagation LSTM	59
Gambar 4.15 Code perhitungan gradien setiap gate	60
Gambar 4.16 Code perhitungan gradien bobot	60
Gambar 4.17 Code perhitungan gradien hidden state	61
Gambar 4.18 Code untuk update bobot	61
Gambar 4.19 Kompilasi model (LSTM manual)	61
Gambar 4.20 Akurasi model LSTM manual (dataset: preprocessed.csv)	62
Gambar 4.21 Akurasi model LSTM manual (dataset: preprocessed_ns.csv)	62
Gambar 4.22 Pembuatan model LSTM menggunakan library Tensorflow Keras	63
Gambar 4.23 Kompilasi Model	63
Gambar 4.24 Pelatihan Model.....	64
Gambar 4.25 Akurasi model LSTM menggunakan library Tensorflow Keras	64
Gambar 4.26 Code evaluasi matriks untuk mencari f-1 score terbaik	65
Gambar 4.27 Code untuk menentukan nilai threshold terbaik	66
Gambar 4.28 Code pembuatan confusion matriks	66
Gambar 4.29 Proses encoding data tanpa Indobert	68
Gambar 4.30 Akurasi model training LSTM tanpa Indobert	68
Gambar 4.31 Akurasi model RNN dengan Indobert	70
Gambar 4.32 Akurasi model RNN tanpa Indobert	70
Gambar 4.33 Halaman utama website pendekripsi berita palsu	72
Gambar 4.34 Halaman hasil klasifikasi berita hoaks	73

Gambar 4. 35 Halaman hasil klasifikasi berita fakta	74
Gambar 4. 36 Grafik Perbandingan Model	74
Gambar 4. 37 Contoh klasifikasi hoaks (benar)	75
Gambar 4. 38 Contoh klasifikasi fakta (benar)	76
Gambar 4. 39 Contoh klasifikasi fakta (Salah)	76
Gambar 4. 40 Contoh klasifikasi Hoaks (Salah)	77
Gambar 4. 41 Pesan error pada model LSTM menggunakan numpy	78
Gambar 4. 42 Proses perhitungan nilai gate LSTM	79
Gambar 4.43 Pesan error pada model LSTM	79

DAFTAR TABEL

Tabel 2.1 Contoh penggunaan Data Cleaning	15
Tabel 2.2 Contoh Tokenization	16
Tabel 2.3 Contoh <i>Stemming</i>	17
Tabel 2.4 Pseudocode LSTM	31
Tabel 2.5 Nilai input, bobot, dan bias	33
Tabel 2.6 Tabel Confusion Matriks (Tharwat, 2021).	39
Tabel 2.7 Penelitian sistem deteksi berita palsu	40
Tabel 4. 1 Contoh teks sebelum di-encode	56
Tabel 4. 2 Hasil confusion matriks	67
Tabel 4. 3 Confusion Matrix Model LSTM tanpa Indobert	69
Tabel 4. 4 Hasil pengujian model LSTM tanpa Indobert	69
Tabel 4. 5 Confusion Matrix Model RNN dengan Indobert	70
Tabel 4. 6 Hasil Pengujian Model RNN dengan Indobert	71
Tabel 4. 7 Confusion Matrix Model RNN tanpa Indobert	71
Tabel 4. 8 Hasil Pengujian Model RNN tanpa Indobert	71
Tabel 4. 9 Rekapitulasi Hasil Pengujian Model	71

DAFTAR PUSTAKA

- Ahdiat, A. (2022). *Kebiasaan Responden dalam Mengonsumsi Berita Online*. Kata Data Media Network.
<https://databoks.katadata.co.id/datapublish/2022/09/15/banyak-warga-ri-belum-terbiasa-menyaring-berita-online>
- Annur, C. M. (2022). *Survei KIC: Mayoritas Masyarakat Indonesia Mengakses Informasi di Media Sosial*. Kata Data Media Network.
<https://databoks.katadata.co.id/datapublish/2022/01/20/survei-kic-majoritas-masyarakat-indonesia-mengakses-informasi-di-media-sosial>
- Annur, C. M. (2023). *Media yang Paling Dipercaya Warga Indonesia pada 2023 [Official Website]*. Katadata Media Network.
<https://databoks.katadata.co.id/datapublish/2023/06/15/inilah-media-yang-paling-dipercaya-warga-indonesia-pada-2023-ada-favoritmu>
- Apuke, O. D., & Omar, B. (2021). User motivation in fake news sharing during the COVID-19 pandemic: An application of the uses and gratification theory. *Online Information Review*, 45(1), 220–239. <https://doi.org/10.1108/OIR-03-2020-0116>
- Arzina, T., Saiduzzaman, Md., Rahman, M. A., Akhter, J., & Rahaman, A. S. Md. M. (2022). Performance Evaluation of Multiple Classifiers for Predicting Fake News. *Journal of Computer and Communications*, 10, 1–21. <https://doi.org/10.4236/jcc.2022.109001>
- Bahad, P., Saxena, P., & Kamal, R. (2019). Fake News Detection using Bi-directional LSTM-Recurrent Neural Network. *Procedia Computer Science*, 165(2019), 74–82. <https://doi.org/10.1016/j.procs.2020.01.072>
- Burkhardt, J. M. (2017). History of fake news. *Library Technology Reports*, 53(8), 1–33.
- Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding* (arXiv:1810.04805). arXiv. <http://arxiv.org/abs/1810.04805>
- Granik, M., & Mesyura, V. (2017). Fake news detection using naive Bayes classifier. *2017 IEEE 1st Ukraine Conference on Electrical and Computer*

- Engineering, UKRC CON 2017 - Proceedings, 900–903.*
<https://doi.org/10.1109/UKRCON.2017.8100379>
- Gumilar, G. (2017). Literasi media: cerdas menggunakan media sosial dalam menanggulangi berita palsu (hoax) oleh siswa sma. *Indonesian Journal of Community Engagement*.
- Hsu, D. (2018). *Multi-period Time Series Modeling with Sparsity via Bayesian Variational Inference* (arXiv:1707.00666). arXiv.
<http://arxiv.org/abs/1707.00666>
- Irena, B., & Setiawan, E. B. (2020). Fake News (Hoax) Identification on Social Media Twitter using Decision Tree C4.5 Method. *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, 4(4), 711–716.
<https://doi.org/10.29207/RESTI.V4I4.2125>
- Jadhav, S. S., & Thepade, S. D. (2019). Fake News Identification and Classification Using DSSM and Improved Recurrent Neural Network Classifier. *Applied Artificial Intelligence*, 33(12), 1058–1068.
<https://doi.org/10.1080/08839514.2019.1661579>
- Jakhotiya, A., Jain, H., Jain, B., Chaniyara, C., Student, B., Professor, A., & -----, M. (2022). Text Pre-Processing Techniques in Natural Language Processing: A Review. *International Research Journal of Engineering and Technology*.
<https://doi.org/10.1109/ACCESS.2017.2672677>
- Kaliyar, R. K., Goswami, A., Narang, P., & Sinha, S. (2020). FNDNet – A deep convolutional neural network for fake news detection. *Cognitive Systems Research*, 61, 32–44. <https://doi.org/10.1016/j.cogsys.2019.12.005>
- Khodra, M. L. (Direktur). (2020, Oktober). *Long Short-Term Memory (LSTM)*.
https://www.youtube.com/watch?v=w7U46SnJnhY&lc=Ugw6S3_6OAecgRO2TJx4AaABA9.9ISfm6pDuPh9IV9IPwTBAu
- Lazer, D. M. J., Baum, M. A., Benkler, Y., Berinsky, A. J., Greenhill, K. M., Menczer, F., Metzger, M. J., Nyhan, B., Pennycook, G., Rothschild, D., Schudson, M., Sloman, S. A., Sunstein, C. R., Thorson, E. A., Watts, D. J., & Zittrain, J. L. (2018). The science of fake news. *Science*, 359(6380), 1094–1096. <https://doi.org/10.1126/science.aao2998>

- Li, Y., & Yang, T. (2018). Word Embedding for Understanding Natural Language: A Survey. *Studies in Big Data*, 26, 83–104. [https://doi.org/10.1007/978-3-319-53817-4_4/COVER](https://doi.org/10.1007/978-3-319-53817-4_4)
- Liu, G., & Guo, J. (2019). Bidirectional LSTM with attention mechanism and convolutional layer for text classification. *Neurocomputing*, 337, 325–338. <https://doi.org/10.1016/j.neucom.2019.01.078>
- Liu, J., & He, D. (2020). Research on The Comment Text Classification based on Transfer Learning. *2020 IEEE 5th Information Technology and Mechatronics Engineering Conference (ITOEC)*, 191–195. <https://doi.org/10.1109/ITOEC49072.2020.9141771>
- Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient Estimation of Word Representations in Vector Space. *1st International Conference on Learning Representations, ICLR 2013 - Workshop Track Proceedings*. <https://arxiv.org/abs/1301.3781v3>
- Mubaraq, M. F., & Maharani, W. (2022). Sentiment Analysis on Twitter Social Media towards Climate Change on Indonesia Using IndoBERT Model. *Jurnal media informatika budidarma*, 6(4), 2426. <https://doi.org/10.30865/mib.v6i4.4570>
- Muhammad, P. F., Kusumaningrum, R., & Wibowo, A. (2021). Sentiment Analysis Using Word2vec and Long Short-Term Memory (LSTM) for Indonesian Hotel Reviews. *Procedia Computer Science*, 179(2020), 728–735. <https://doi.org/10.1016/j.procs.2021.01.061>
- Nugroho, S. E. (2017). *Kementerian Komunikasi dan Informatika*. Kementerian Komunikasi dan Informatika.
- Nurdin, A., Anggo, B., Aji, S., Bustamin, A., & Abidin, Z. (2020). Perbandingan kinerja word embedding word2vec, glove, dan fasttext pada klasifikasi teks. *Jurnal Tekno Kompak*, 14(2), 74–79. <https://doi.org/10.33365/JTK.V14I2.732>
- Olah, C. (2015). *Understanding LSTM Networks*. Colah's blog. <http://colah.github.io/posts/2015-08-Understanding-LSTMs/>

- Pennington, J., Socher, R., & Manning, C. D. (2014). GloVe: Global Vectors for Word Representation. *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, 1532–1543. <http://nlp>.
- Putra, H. K., Bijaksana, M. A., & Romadhony, A. (t.t.). *Deteksi Penggunaan Kalimat Abusive Pada Teks Bahasa Indonesia Menggunakan Metode IndoBERT*.
- Raschka, S., & Mirjalili, V. (2019). Python Machine Learning: Machine Learning and Deep Learning with Python. Dalam J. Malysiak (Ed.), *Packt Publishing Ltd.* Packt Publishing Ltd.
- Santoso, A. S. (2017). *Kementerian Komunikasi dan Informatika*. Ada 800.000 Situs Penyebar Hoax di Indonesia. https://www.kominfo.go.id/content/detail/12008/ada-800000-situs-penyebar-hoax-di%20indonesia/0/sorotan_media
- Selivanov, D. (2022). *GloVe Word Embeddings*. GloVe Word Embeddings. <https://cran.r-project.org/web/packages/text2vec/vignettes/glove.html>
- Sen, S., Sugiarto, D., & Rochman, A. (2020). *Komparasi Metode Multilayer Perceptron (MLP) dan Long Short Term Memory (LSTM) dalam Peramalan Harga Beras*. 12(1), 35–41.
- Sherstinsky, A. (2020). Fundamentals of Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) network. *Physica D: Nonlinear Phenomena*, 404, 132306. <https://doi.org/10.1016/j.physd.2019.132306>
- Tabatabaei, S. A., Klein, J., & Hoogendoorn, M. (2020). Estimating the F1 Score for Learning from Positive and Unlabeled Examples. *Lecture Notes in Computer Science*, 150–161.
- Tala, F. (2003). *A Study of Stemming Effects on Information Retrieval in Bahasa Indonesia*.
- Tharwat, A. (2018). Classification assessment methods. *Applied Computing and Informatics*, 17(1), 168–192. <https://doi.org/10.1016/j.aci.2018.08.003>
- Tian, C., Ma, J., Zhang, C., & Zhan, P. (2018). A Deep Neural Network Model for Short-Term Load Forecast Based on Long Short-Term Memory Network and Convolutional Neural Network. *Energies*, 11(12), 3493. <https://doi.org/10.3390/en11123493>

- Utami, M. P., Nurhayati, O. D., & Warsito, B. (2020, Desember 14). Hoax Information Detection System Using Apriori Algorithm and Random Forest Algorithm in Twitter. *6th International Conference on Interactive Digital Media, ICIDM 2020*. <https://doi.org/10.1109/ICIDM51048.2020.9339648>
- Vuković, M., Pripužić, K., & Belani, H. (2009). An intelligent automatic hoax detection system. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 5711 LNAI(PART 1), 318–325. https://doi.org/10.1007/978-3-642-04595-0_39/COVER