

**RANCANG BANGUN APLIKASI *FAKE NEWS DETECTOR* MENGGUNAKAN
METODE *LONG SHORT TERM MEMORY***

SKRIPSI

Diajukan untuk memenuhi bagian dari
syarat memperoleh gelar Sarjana Komputer
pada Departemen Pendidikan Ilmu Komputer
Program Studi Ilmu Komputer



oleh

Bisma Wahyu Anaafie

1505066

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

2019

**RANCANG BANGUN APLIKASI *FAKE NEWS DETECTOR* MENGGUNAKAN
METODE *LONG SHORT TERM MEMORY***

oleh
Bisma Wahyu Anaafie

Sebuah skripsi yang diajukan untuk memenuhi salah satu syarat memperoleh gelar
Sarjana Komputer pada Fakultas Pendidikan Matematika dan Ilmu Pengetahuan Alam

© Bisma Wahyu Anaafie
Universitas Pendidikan Indonesia
September 2019

Hak Cipta dilindungi undang-undang.
Skripsi ini tidak boleh diperbanyak seluruhnya atau sebagian,
dengan dicetak ulang, difoto kopi, atau cara lainnya tanpa ijin dari penulis.

BISMA WAHYU ANAAFIE

RANCANG BANGUNG APLIKASI *FAKE NEWS* DETECTOR MENGGUNAKAN METODE
LONG SHORT-TERM MEMORY

disetujui dan disahkan oleh pembimbing:

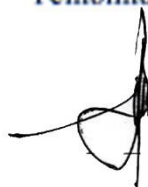
Pembimbing I



Dr. Yudi Wibisono, MT.

NIP. 197507072003121003

Pembimbing II



Yaya Wihardi, S.Kom., M.Kom.

NIP. 198903252015041001

Mengetahui,

Ketua Departemen Pendidikan Ilmu Komputer



Lala Septem Riza, M.T., Ph.D.

NIP. 197809262008121001

RANCANG BANGUN APLIKASI *FAKE NEWS DETECTOR* MENGUNAKAN METODE *LONG SHORT TERM MEMORY*

Bisma Wahyu Anaafie

Departemen Pendidikan Ilmu Komputer

Fakultas Pendidikan Matematika dan Ilmu Pengetahuan Alam

Universitas Pendidikan Indonesia

bismawahyu@student.upi.edu

ABSTRAK

News atau berita merupakan sebuah uraian tentang fakta dan atau pendapat yang mengandung nilai berita dan yang sudah disajikan melalui media masa secara periodik. Di Indonesia berita menjadi suatu hal yang sangat penting mengetahui beragam informasi terkini terutama seputar informasi dalam negeri karena wilayah Indonesia yang sangat luas. Diantara banyaknya informasi yang beredar di tengah masyarakat Indonesia, juga beredar informasi atau berita palsu yang biasa disebut *fake news*. *Fake news* menyebabkan keresahan dan *miss information* serta cenderung menyesatkan bagi para pembacanya. Seiring dengan semakin banyaknya berita palsu, banyak peneliti yang membuat sistem pintar atau *artificial intelligence* untuk mendeteksi dan menyaring berita palsu yang beredar dengan menggunakan algoritma seperti *naïve bayes*, *support vector machine*, bahkan *deep learning*. Pada penelitian ini, algoritma yang digunakan merupakan algoritman *Long Short-Term Memory* (LSTM) yang dikategorikan sebagai *deep learning*. Dataset diambil dari portal berita Indonesia sebagai sumber berita asli dan situs penampungan berita palsu sebagai sumber berita palsu, metode *case folding* untuk membuat standar teks menjadi *lower case* secara keseluruhan, *word embedding* untuk merepresentasikan data teks menjadi vektor, dalam penelitian ini digunakan tiga metode representasi teks, yaitu (1)*CountVectorizer*, (2)*TF-IDF*, (3)*Word2Vec*. Pelatihan data menggunakan metode *Long Short-Term Memory* (LSTM) dengan library Keras. Hasilnya, sistem dapat memprediksi berita dengan akurasi 52.22%.

DESIGN OF FAKE NEWS DETECTOR APPLICATION USING LONG SHORT-TERM MEMORY METHOD

Bisma Wahyu Anaafie

*Computer Science Education Department
Faculty of Mathematics and Science Education
Indonesia University of Education*

bismawahyu@student.upi.edu

ABSTRACT

News is a description of facts and / or opinions that contain news value and which have been presented through mass media periodically. In Indonesia the news becomes very important to know a variety of current information, especially about domestic information because of Indonesia's vast territory. Among the many information circulating in the middle of Indonesian society, also circulated false information or news which is commonly called fake news. Fake news causes anxiety and miss information and tends to be misleading for its readers. Along with the increasing number of fake news, many researchers are creating intelligent systems or artificial intelligence to detect and filter out fake news that is circulating by using algorithms such as naïve bayes, support vector machines, even deep learning. In this study, the algorithm used is the Long Short-Term Memory (LSTM) algorithm which is categorized as deep learning. The dataset is taken from the Indonesian news portal as an original news source and a fake news collection site as a fake news source, a case folding method to make the text standard a lower case as a whole, word embedding to represent text data into vectors, in this study three teks representation methods are used, namely (1) CountVectorizer, (2) TF-IDF, (3) Word2Vec. Training data using the Long Short-Term Memory (LSTM) method with the Hard library. The result, the system can predict news with an accuracy of 52.22%.

DAFTAR ISI

ABSTRAK	i
ABSTRACT	ii
DAFTAR ISI.....	iii
DAFTAR TABEL.....	vi
DAFTAR GAMBAR	viii
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang Masalah.....	1
1.2 Rumusan Masalah	2
1.3 Tujuan Penelitian.....	3
1.4 Manfaat Penelitian.....	3
1.5 Batasan Masalah.....	3
1.6 Sistematika Penulisan.....	4
BAB II KAJIAN PUSTAKA.....	6
2.1 Berita	6
2.2 Berita Palsu (<i>Fake News</i>)	7
2.3 Penelitian Terkait	9
2.6 <i>Word Embedding</i>	12
2.6.1. <i>Word2Vec</i>	12
2.7 <i>Naïve Bayes</i>	13
2.8 <i>K Nearest Neighbor (KNN)</i>	14
2.9 <i>Decision Tree</i>	14
2.10 <i>Support Vector Machine</i>	15

2.11	<i>Random Forest</i>	16
2.12	<i>Artificial Neural Network</i>	17
2.11.1.	Loss Function	20
2.11.2.	Hyperparameter.....	21
2.13	Deep Learning.....	21
2.14	Recurrent Neural Network.....	22
2.13.1.	Pemodelan Bahasa	23
2.15	<i>Long Short-Term Memory</i>	24
2.14.1	Arsitektur Long Short-Term Memory.....	25
BAB III METODOLOGI PENELITIAN		28
3.1	Desain Penelitian	28
3.2	Alat dan Bahan Penelitian	29
3.2.1.	Alat Penelitian.....	29
3.2.2.	Bahan Penelitian.....	29
3.3	Metode Penelitian.....	31
3.3.1.	Metode Pengumpulan Data.....	31
3.3.2.	Metode Pengembangan Perangkat Lunak.....	31
BAB IV PEMBAHASAN.....		34
4.1.	Data	34
4.1.1.	Praproses Data.....	35
4.2.	Eksperimen Naïve Bayes.....	37
4.3.	Eksperimen <i>K Nearest Neighbor</i> (KNN)	38
4.4.	Eksperimen <i>Decision Tree</i>	38
4.5.	Eksperimen <i>Support Vector Machine</i>	39

4.6.	Eksperimen <i>Random Forest</i>	39
4.7.	Eksperimen <i>Long Short-Term Memory</i>	40
4.8.	Validasi Model	43
4.9.	Analisis Hasil	44
BAB V KESIMPULAN DAN SARAN.....		50
5.1	Kesimpulan.....	50
5.2	Saran.....	51
DAFTAR PUSTAKA		52

DAFTAR TABEL

Tabel 2. 1. Contoh proses representasi teks menggunakan CountVectorizer	10
Tabel 2. 2. Beberapa parameter yang digunakan dalam algoritma random forest.....	17
Tabel 4. 1. Sample dataset berita yang digunakan	33
Tabel 4. 2. Hasil praproses dengan case folding	34
Tabel 4. 3. Contoh hasil representasi teks menggunakan CountVectorizer.....	35
Tabel 4. 4. Sample hasil representasi kata menggunakan metode TF-IDF.....	36
Tabel 4. 5. Tabel hasil eksperimen menggunakan metode Naïve Bayes	36
Tabel 4. 6. Hasil Eksperimen menggunakan KNN Classifier.....	37
Tabel 4. 7. Hasil eksperimen menggunakan Decision Tree Classifier	37
Tabel 4. 8. Hasil eksperimen menggunakan Support Vector Machine Classifier	38
Tabel 4. 9. Hasil eksperimen menggunakan Random Forest.....	38
Tabel 4. 10. Hasil eksperimen menggunakan Random Forest + TF-IDF	39
Tabel 4. 11. Proses Tuning Parameter pada Eksperimen Count Vectorizer Text Representation.....	39
Tabel 4. 12. Proses Tuning Parameter Eksperimen TF-IDF Text Representation.....	40
Tabel 4. 13. Proses Tuning Parameter pada Eksperimen Word2Vec Word Embedding	41
Tabel 4. 14. Hasil Eksperimen menggunakan algoritma Random Forest + TF-IDF.....	43
Tabel 4. 15. Hasil validasi setiap kelas menggunakan algoritma random forest	43
Tabel 4. 16. Confusion Matrix hasil testing model klasifikasi Random Forest.....	44
Tabel 4. 17. Tiga kata dengan frekuensi tertinggi di kelas berita asli.....	44
Tabel 4. 18. Tiga kata dengan frekuensi tertinggi di kelas berita palsu.....	44
Tabel 4. 19. Contoh penggunaan kata “orang” di tiap kelas.....	45
Tabel 4. 20. Hasil Evaluasi model klasifikasi menggunakan algoritma LSTM+Word2Vec	45
Tabel 4. 21. Hasil validasi setiap kelas	46

Tabel 4. 22. Confusion Matrix hasil testing model klasifikasi model LSTM+Word2Vec
..... 46

DAFTAR GAMBAR

Gambar 2. 1. Ilustrasi relasi antar kata dengan menggunakan metode Word2Vec	13
Gambar 2. 2. Ilustrasi algoritma Decision Tree	14
Gambar 2. 3. Ilustrasi struktur algoritma random forest.....	16
Gambar 2. 4. Ilustrasi bentuk jaringan pada ANN.....	18
Gambar 2. 5. Fungsi Threshold.....	19
Gambar 2. 6. Fungsi Sigmoid	19
Gambar 2. 7. Fungsi Identitas	19
Gambar 2. 8 Ilustrasi arsitektur RNN	23
Gambar 2. 9. Ilustrasi arsitektur LSTM.....	26
Gambar 3.1. Desain Penelitian.....	28
Gambar 4. 1. Contoh berita palsu yang bersifat manipulasi	47
Gambar 4. 2. Contoh berita palsu yang diprediksi berita asli.....	48
Gambar 4. 3. Sumber Berita Asli yang di Manipulasi	48

DAFTAR PUSTAKA

- Ahmed, H., Traore, I., & Saad, S. (2017). Detecting opinion spams and fake news using text classification, (September), 1–15. <https://doi.org/10.1002/spy2.9>
- Bogers, T. (2007). Comparing and Evaluating Information Retrieval Algorithms for News Recommendation, 141–144.
- Conroy, N. J., Rubin, V. L., & Chen, Y. (2015). Automatic Deception Detection : Methods for Finding Fake News. *Automatic Deception Detection : Methods for Finding Fake News*.
- Dervin, B., & Editor, G. (2009). Information : An Answer for Every Question ? A Solution for Every Problem ?, (February 2015), 37–41. <https://doi.org/10.1080/08838157609386402>
- Freeman, D. M., & Hwa, T. (n.d.). Detecting Clusters of Fake Accounts in Online Social Networks Categories and Subject Descriptors, 91–101.
- Khadjeh, A., Aghabozorgi, S., Ying, T., Chek, D., & Ngo, L. (2015). Expert Systems with Applications Text mining of news-headlines for FOREX market prediction : A Multi-layer Dimension Reduction Algorithm with semantics and sentiment. *EXPERT SYSTEMS WITH APPLICATIONS*, 42(1), 306–324. <https://doi.org/10.1016/j.eswa.2014.08.004>
- Mei, Q. (2005). Discovering Evolutionary Theme Patterns from Text - An Exploration of Temporal Text Mining, 198–207.
- Meservy, T. O., Jensen, M. L., Kruse, J., Burgoon, J. K., Jay, F., Twitchell, D. P., ... Metaxas, D. N. (n.d.). through Automatic , of Nonverbal Behavior.
- Methods, P., Chakraborty, G., & Pagolu, M. (n.d.). Text Mining and Analysis.
- Of, C., & Acm, T. H. E. (2000). Machine earnin D emand.
- Phuvipadawat, S., & Murata, T. (2010). Breaking News Detection and Tracking in Twitter. <https://doi.org/10.1109/WI-IAT.2010.205>
- Rubin, V. L., Conroy, N. J., & Chen, Y. (2015). Towards News Verification : Deception Detection Methods for News Discourse.
- Tseng, Y., Lin, C., & Lin, Y. (2007). Text mining techniques for patent analysis, 43, 1216–1247. <https://doi.org/10.1016/j.ipm.2006.11.011>

- Feldman, R., and Sanger, J. (2006). *The Text Mining Handbook*. New York: Cambridge University Press. ISBN 978-0-521-83657-9
- Lazonder, A. W., Biemans, H. J. a, & Wopereis, I. G. J. H. (2000). Differences between novice and experienced users in search information on the World Wide Web. [https://doi.org/10.1002/\(sici\)1097-4571\(2000\)51:6<576::aid-asi9>3.0.co;2-7](https://doi.org/10.1002/(sici)1097-4571(2000)51:6<576::aid-asi9>3.0.co;2-7)
- Liddy, E.D., 2001, Natural Language Processing, In Encyclopedia of Library and Information Science, 2nd Edition, Marcel Decker Inc, NY, USA
- LeCun, Y., Bengio, Y., & Hinton, G. (2015). *Deep learning*. *Nature*, 521(7553), 436–444.
- Deng, L. (2014). Deep Learning: Methods and Applications. Foundations and Trends® in Signal Processing, 7(3-4), 197–387. doi:10.1561/20000000039
- Rosasco, L.; De Vito, E. D.; Caponnetto, A.; Piana, M.; Verri, A. (2004). "Are Loss Functions All the Same?" (PDF). *Neural Computation*. 16 (5): 1063–1076. CiteSeerX 10.1.1.109.6786
- Rasywir, E., & Purwarianti, A. (2016). Eksperimen pada Sistem Klasifikasi Berita Hoax Berbahasa Indonesia Berbasis Pembelajaran Mesin. *Jurnal Cybermatika*, 3(2).
- Ishak, A., Chen, Y. Y., & Yong, S. P. (2012, June). Distance-based hoax detection system. In *2012 International Conference on Computer & Information Science (ICIS)* (Vol. 1, pp. 215-220). IEEE.
- Vuković, M., Pripuzić, K., & Belani, H. (2009, September). An intelligent automatic hoax detection system. In *International Conference on Knowledge-Based and Intelligent Information and Engineering Systems* (pp. 318-325). Springer, Berlin, Heidelberg.
- Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, 9(8), 1735-1780.

- Ho, Tin Kam (1995). [Random Decision Forests](#) (PDF). *Proceedings of the 3rd International Conference on Document Analysis and Recognition, Montreal, QC, 14–16 August 1995*. hlm. 278–282
- Youngjoong Ko (2012). "A study of term weighting schemes using class information for text classification". [SIGIR'12](#). [ACM](#).
- Firth, J.R. (1957). "A synopsis of linguistic theory 1930-1955". *Studies in Linguistic Analysis*: 1–32. Reprinted in F.R. Palmer, ed. (1968). *Selected Papers of J.R. Firth 1952-1959*. London: Longman.
- Neelakantan, Arvind; Shankar, Jeevan; Passos, Alexandre; McCallum, Andrew (2014). "[Efficient Non-parametric Estimation of Multiple Embeddings per Word in Vector Space](#)". *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Stroudsburg, PA, USA: Association for Computational Linguistics. [doi:10.3115/v1/d14-1113](#)
- Qureshi, M. Atif; Greene, Derek (2018-06-04). "EVE: explainable vector based embedding technique using Wikipedia". *Journal of Intelligent Information Systems*. [arXiv:1702.06891](#). [doi:10.1007/s10844-018-0511-x](#). [ISSN 0925 9902](#).
- Mayer, D. G.; Butler, D.G. (1993), "Statistical validation", [Ecological Modelling](#), **68**: 21–32, [doi:10.1016/0304-3800\(93\)90105-2](#).
- Powers, David M. W. (2012). "[The Problem with Kappa](#)". *Conference of the European Chapter of the Association for Computational Linguistics (EACL2012) Joint ROBUS-UNSUP Workshop*.