

Identifikasi Berita *Hoax* dengan Recurrent Neural Network

Kevin Perdana¹, Zulfachmi², Dwi Nurul Huda³

^{1,2,3} STT Indonesia Tanjungpinang
Jln. Pompa Air No. 28 Tanjungpinang Kepulauan Riau Indonesia

¹kevin@sttindonesia.ac.id

²fachmi@sttindonesia.ac.id

³dwi.nurulhuda@gmail.com

Intisari— Berita *hoax* merupakan masalah universal yang dapat merubah sudut pandang seseorang. Dampaknya dapat menimbulkan ketakutan, memunculkan ide-ide rasis, dan mengarah pada penindasan dan kekerasan terhadap orang yang tidak bersalah. Masalah ini sudah sejak zaman dahulu walaupun belum ada akses informasi ke seluruh dunia. Saat ini terdapat 2,5 triliun byte data dan terus meningkat sehingga membawa ke masalah berikutnya yaitu menganalisis dan mengidentifikasi berita *hoax* dan berita nyata agar lebih cepat. Solusi yang ideal adalah dengan algoritma Deep Learning dan penulis memilih metode Recurrent Neural Network (RNN) dan variannya Long Short-Term Memory (LSTM). Akurasi yang didapat sebesar 99%, lebih besar daripada metode Machine Learning seperti Rocchio ataupun Multinomial Naive Bayes.

Kata kunci— Berita, *Hoax*, Machine Learning, RNN, LSTM.

Abstract— *The problem of hoax news is something universal that can change someone perspective. The impact can generate fear, racism ideas, and lead to oppression and violence against innocent people. This problem has existed since time immemorial, even though there is no access to information throughout the world. Currently there are 2.5 trillion bytes of data and it increasing, leading to the next problem, i.e. faster at analyzing and identifying hoax news and real news. The ideal solution is using Deep Learning algorithm and the author chooses Recurrent Neural Network (RNN) method and it's particular method Long Short-Term Memory (LSTM). The accuracy obtained is 99%, greater than Machine Learning methods such as Rocchio or Multinomial Naive Bayes.*

Keywords— News, Hoax, Machine Learning, RNN, LSTM.

I. PENDAHULUAN

Fenomena berita/informasi palsu atau berita *hoax* sangat mempengaruhi kehidupan sosial kita [1]. Karena informasi membentuk cara kita memandang alam semesta, berita *hoax* dan fakta menyesatkan lainnya dapat memberikan sudut pandang yang berbeda. Penelitian di bidang intelijen menunjukkan bahwa rumor yang menyebar di media sosial meninggalkan dampak yang bertahan lama pada orang yang kurang pintar dan membuat mereka tidak mengambil keputusan yang tepat. Berita *hoax* digunakan untuk membangun ketakutan orang, memunculkan ide-ide rasis, dan mengarah pada penindasan dan kekerasan terhadap orang yang tidak bersalah [2].

Sejak dahulu berita *hoax* merupakan masalah universal. Walaupun di masa lalu tidak ada akses informasi ke seluruh dunia, pendeteksian berita *hoax* relatif sulit dan tidak efisien dari segi biaya. Namun saat ini sangat mudah dan layak untuk mengidentifikasi apakah berita itu *hoax* atau nyata. Yang membawa kita ke masalah berikutnya dalam menganalisis dan mengidentifikasi berita *hoax* dari berita nyata dalam Gigabytes informasi. Ini mungkin fakta yang membingungkan bahwa ada 2,5 triliun byte data yang dibuat setiap hari dengan kecepatan kita saat ini dan kecepatan ini tidak menurun dari waktu ke waktu [3].

Penelitian yang berkaitan dengan *hoax* adalah yang dilakukan Aulia Afriza dan Julio Adisantoso dengan judul Metode Klasifikasi Rocchio untuk Analisis *Hoax*. Nilai akurasi

yang didapat dengan Metode Rocchio sebesar 83,501% sedangkan Multinomial Naive Bayes sebesar 65,835% [4].

Kemudian penelitian yang dilakukan Putu Kussa Laksana Utama dengan judul Identifikasi *Hoax* pada Media Sosial dengan Pendekatan Machine Learning, disimpulkan bahwa solusi yang ideal untuk identifikasi konten *hoax* adalah dengan algoritma Deep Learning agar proses klasifikasi dapat dilakukan dengan cepat [5].

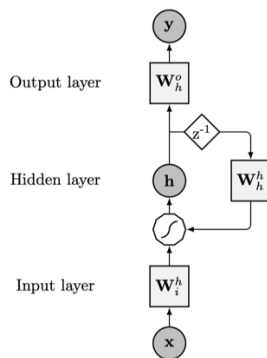
II. METODOLOGI PENELITIAN

Algoritma yang digunakan penulis adalah salah satu dari sekian banyak dalam bidang Deep Learning yaitu Recurrent Neural Network (RNN) dan variannya Long Short-Term Memory (LSTM).

Recurrent Neural Network (RNN)

Recurrent Neural Network (RNN) adalah kelas jaringan saraf tiruan dimana koneksi antar node membentuk garis yang diarahkan di sepanjang urutan temporal. RNN dapat menggunakan keadaan internal (memori) untuk memproses inputan variabel sekuensial [6]. Dengan kata lain cara perhitungan *state* (keadaan) berjalan dengan menerapkan fungsi ke *state* sebelumnya dan *state* inputan secara rekursif [7]. RNN telah berhasil digunakan untuk berbagai tugas seperti pemodelan bahasa, *word embedding* yaitu proses konversi kata yang berupa karakter ke dalam bentuk *vector*, pengenalan tulisan tangan, dan pengenalan ucapan [8].

Arsitektur RNN ditunjukkan pada gambar di bawah ini.



Gambar 1. Arsitektur RNN [7]

Lingkaran mewakili *node*, x = input, h = hidden, dan y = output. Simbol di dalam kotak, w_i^h , w_h^h dan w_h^o adalah matriks yang masing-masing mewakili bobot input, hidden dan output. Poligon mewakili transformasi non-linear yang dilakukan oleh neuron dan z^{-1} adalah *unit delay operator* [7].

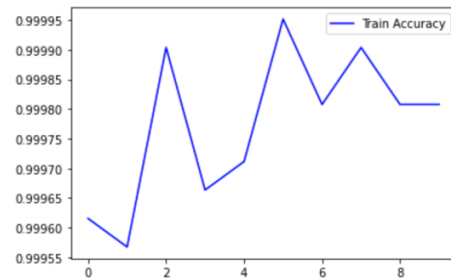
Long Short-Term Memory (LSTM)

Long Short-Term Memory (LSTM) adalah salah satu arsitektur RNN yang banyak digunakan pada masalah Deep Learning [9] yang melibatkan teks [10]. LSTM dapat mengingat informasi jangka panjang. Ide dari LSTM adalah membuatnya jalur yang menghubungkan konteks lama ke konteks baru yang disebut juga *cell state*, *memory cell* atau jalur memori. Dengan adanya jalur tersebut, suatu nilai pada konteks yang lama akan dengan mudah dihubungkan ke konteks yang baru jika diperlukan, dengan sedikit sekali modifikasi [11]. Sebagai model RNN khusus, LSTM memiliki ekspresi fitur yang lebih detail daripada model RNN tradisional [12].

III. HASIL DAN PEMBAHASAN

Penelitian ini terdiri dari beberapa langkah yaitu :

1. Koleksi judul berita Indonesia sebanyak 250 data.
2. Memberi label *hoax* dan valid secara manual dan disimpan dalam file csv.
3. Kemudian memisahkan file csv menjadi 2 file. 210 data sebagai data *training* dan 40 data sebagai data *test*.
4. Mengubah kalimat pada berita menjadi Bahasa Inggris.
5. Menghilangkan *stopwords*. *Stopwords* artinya kata-kata tidak berarti yang memiliki daya diskriminasi rendah [13] atau kata-kata dengan nilai informatif rendah [14].
6. Stemming. Secara definisi adalah mengurangi berbagai bentuk tata bahasa dari sebuah kata menjadi kata dasar [15].
7. Memulai training dengan RNN.



Gambar 2. Grafik Proses Training

Berdasarkan grafik diatas dapat dilihat klasifikasi berita hoax menggunakan RNN menghasilkan nilai akurasi sebesar 99%.

IV. KESIMPULAN

Klasifikasi dengan Recurrent Neural Network menghasilkan akurasi dengan sangat baik yaitu 99%. Nilai akurasi ini tentu lebih baik ketimbang menggunakan metode Machine Learning lain seperti Rocchio ataupun Multinomial Naive Bayes. Dan proses training dilakukan sangat cepat, tentunya dapat diterapkan untuk menangani data dalam jumlah besar (Gigabytes).

Penelitian selanjutnya diharapkan melakukan analisa tidak hanya judul beritanya saja melainkan isi beritanya juga.

REFERENSI

- [1] S. Vinit Bhoir, "An efficient fake news detector," *2020 Int. Conf. Comput. Commun. Informatics, ICCCI 2020*, 2020, doi: 10.1109/ICCCI48352.2020.9104177.
- [2] M. G. Hussain, M. Rashidul Hasan, M. Rahman, J. Protim, and S. Al Hasan, "Detection of Bangla Fake News using MNB and SVM Classifier," *Proc. - 2020 Int. Conf. Comput. Electron. Commun. Eng. iCCECE 2020*, pp. 81–85, 2020, doi: 10.1109/iCCECE49321.2020.9231167.
- [3] S. Sharma and D. K. Sharma, "Fake News Detection: A long way to go," *2019 4th Int. Conf. Inf. Syst. Comput. Networks, ISCON 2019*, pp. 816–821, 2019, doi: 10.1109/ISCON47742.2019.9036221.
- [4] A. Afriza and J. Adisantoso, "Metode Klasifikasi Rocchio untuk Analisis Hoax," *J. Ilmu Komput. dan Agri-Informatika*, vol. 5, no. 1, p. 1, 2018, doi: 10.29244/jika.5.1.1-10.
- [5] P. K. Laksana Utama, "Identifikasi Hoax pada Media Sosial dengan Pendekatan Machine Learning," *Widya Duta J. Ilm. Ilmu Agama dan Ilmu Sos. Budaya*, vol. 13, no. 1, p. 69, 2018, doi: 10.25078/wd.v13i1.436.
- [6] S. Naderian, "A Novel Hybrid Deep Learning Approach for Non-Intrusive Load Monitoring of Residential Appliance Based on Long Short Term Memory and Convolutional Neural Networks."
- [7] F. M. Bianchi, E. Maiorino, M. C. Kampffmeyer, A. Rizzi, and R. Jenssen, *An overview and comparative analysis of Recurrent Neural Networks for Short Term Load Forecasting*, no. November. 2017.
- [8] R. Pascanu, C. Gulcehre, K. Cho, and Y. Bengio, "How to construct deep recurrent neural networks," *2nd Int. Conf. Learn. Represent. ICLR 2014 - Conf. Track Proc.*, no. December, 2014.
- [9] A. S. Talita and A. Wiguna, "Implementasi Algoritma Long Short-Term Memory (LSTM) Untuk Mendeteksi Ujaran Kebencian (Hate Speech) Pada Kasus Pilpres 2019," *MATRIK J. Manajemen, Tek. Inform. dan Rekayasa Komput.*, vol. 19, no. 1, pp. 37–44, 2019, doi: 10.30812/matrik.v19i1.495.
- [10] O. Ajao, D. Bhowmik, and S. Zargari, "Fake News Identification on Twitter with Hybrid CNN and RNN Models," *arXiv*, no. June, 2018.
- [11] J. Pardede and R. G. Ibrahim, "Implementasi Long Short-Term Memory untuk Identifikasi Berita Hoax Berbahasa Inggris pada Media Sosial," *J. Comput. Sci. Informatics Eng.*, vol. 4, no. 2, pp. 179–187, 2020, doi: 10.29303/jcosine.v4i2.361.

- [12] C. Pan, J. Tan, D. Feng, and Y. Li, "Very Short-Term Solar Generation Forecasting Based on LSTM with Temporal Attention Mechanism," *2019 IEEE 5th Int. Conf. Comput. Commun. ICC3 2019*, pp. 267–271, 2019, doi: 10.1109/ICCC47050.2019.9064298.
- [13] H. Saif, M. Fernandez, Y. He, and H. Alani, "On stopwords, filtering and data sparsity for sentiment analysis of twitter," *Proc. 9th Int. Conf. Lang. Resour. Eval. Lr. 2014*, no. i, pp. 810–817, 2014.
- [14] H. Saif, M. Fernandez, and H. Alani, "Automatic stopword generation using contextual semantics for sentiment analysis of Twitter," *CEUR Workshop Proc.*, vol. 1272, no. i, pp. 281–284, 2014.
- [15] Anjali, G. Jivani, and M. Anjali, "A Comparative Study of Stemming Algorithms," *October*, vol. 2, no. 2004, pp. 1930–1938, 2007.