**BAB I**

**PENDAHULUAN**

* 1. **Latar Belakang**

Penyebaran berita *hoax* saat ini sedang marak terjadi, bahkan sudah masuk fase yang cukup mengkhawatirkan. Menteri Komunikasi dan Informatika (Menkominfo), Rudiantara mengatakan tidak kurang dari 900 ribu situs penyebar informasi *hoax* (Fajri, 2019). Penyebaran berita *hoax* ini bertujuan untuk menggiring opini para pembaca sehingga memiliki persepsi yang sama dengan pelaku. Apabila penyebaran berita *hoax* dibiarkan akan berdampak buruk bagi masyarakat seperti pengaburan fakta, hilangnya kepercayaan masyarakat hingga dapat mengakibatkan perpecahan antar masyarakat.

Media yang sering digunakan untuk melakukan penyebaran berita *hoax* seperti media cetak, email, situs web dan media sosial. Dari beberapa media penyebaran *hoax* tersebut, menurut hasil survei Masyarakat Telematika Indonesia (Mastel) menunjukkan bahwa media sosial menjadi sumber utama peredaran berita *hoax* (Librianty, 2017). Hal tersebut karena tingginya pengguna media sosial di Indonesia yang mencapai 150 juta orang atau sebesar 56% dari total populasi dimana jumlah tersebut naik 20% (Wearesocial, 2019). Ada beberapa jenis berita *hoax* yang sering diterima masyarakat seperti isu di bidang sosial politik, SARA, kesehatan, IPTEK, bencana alam, dan lain-lain. Banyaknya berita *hoax* yang tersebar membuat sulitnya masyarakat dalam memilah dan membedakan mana berita *hoax* dan tidak. Ada beberapa masyarakat yang kritis terhadap berita yang tersebar, lalu melakukan pengecekan apakah berita tersebut *hoax* atau tidak. Namun, tidak sedikit pula masyarakat yang langsung percaya terhadap berita yang disampaikan. Maka dari itu, perlu dikembangkan sebuah sistem cerdas yang dapat digunakan untuk mengklasifikasikan suatu konten berita sehingga dapat diketahui apakah berita tersebut *hoax* atau bukan.

Penelitian yang berkaitan dengan *hoax* pernah dilakukan menggunakan metode *naïve bayes* dengan tingkat akurasi 80% untuk mengklasifikasikan 3 kategori yaitu berita palsu (*hoax*), berita asli (*fact*) dan berita yang tidak berindikasi (Tanjung, 2018). Selain itu, ada penelitian lain yang membandingkan antara algoritma *rocchio* dengan *multinomial naïve bayes* dengan akurasi yang didapatkan sebesar 83,501% untuk *rocchio* dan 65,835% untuk *multinomial naïve bayes* (Afriza & Adisantoso, 2018). Algoritma *support vector machine* pun pernah digunakan untuk klasifikasi artikel *hoax* dengan tingkat akurasi yang didapatkan sebesar 95,8333% (Maulina & Sagara, 2018). Ada beberapa metode lain yang pernah digunakan seperti *gradient descent backpropagation* dengan tingkat akurasi 72,04% sedangan *lavenberg-marquad backpropagation* memiliki akurasi 72,19% dalam klasifikasi pengguna media sosial (Lhaksmana, Nhita, & Budhiarto, 2017). Penerapan *deep learning* juga pernah dilakukan dalam klasifikasi sentimen analisis *tweet* dengan menggunakan metode *deep belief network* dengan tingkat akurasi sebesar 93,31% (Zulfa & Winarko, 2017).

Ada tantangan yang harus dihadapi dalam proses klasifikasi berita *hoax* yaitu banyaknya berita *hoax* yang memiliki tingkat kemiripan yang sangat tinggi dengan berita aslinya. Bahkan dalam beberapa kasus, penambahan atau perubahan beberapa kata dari berita asli akan merubah maksud dan tujuan dari berita tersebut. Seperti kendala dari penggunaan metode *naïve bayes*, dimana kemiripan data membuat sulitnya menemukan pola berita *hoax* sehingga mempengaruhi tingkat error dan akurasinya (Tanjung, 2018). Selain itu, banyaknya data yang digunakan membuat waktu yang diperlukan untuk melakukan *preprocessing* dan *training* semakin lama. Untuk meminimalisisr hal tersebut, dapat dilakukan dengan menggunakan metode *deep learning*.

Kelebihan dari penggunaan metode *deep learning* tersebut karena memiliki algoritma dengan struktur dan jumlah jaringan syaraf yang sangat banyak sebagai jaringan tersembunyi (*hidden layer*) yang dapat digunakan untuk berbagai hal, salah satunya klasifikasi teks. Metode *deep learning* yang dapat digunakan untuk klasifikasi teks yaitu *Convolutional Neural Network (CNN), Recurrent Neural Network (RNN),* dan *Hierarchical Attention Network (HAN)*. Ketiga metode tersebutmemiliki layer hirarki banyak/bertumpuk-tumpuk sehingga kendala yang berkaitan dengan kemiripan antar data dapat diminimalisir dan hasil yang didapat menjadi lebih akurat. Kelebihan lain yang dimiliki metode-metode tersebut yaitu menghilangkan ekstraksi fitur secara manual, tidak memerlukan informasi terhadap data yang akan dipelajarinya, dan dapat dikombinasikan dengan *Graphics Processing Unit* (GPU) untuk mempercepat proses *preprocessing* maupun *training*. Algoritma dari *deep learning* dapat melakukan *tuning* (penyetelan) dan pemilihan model yang paling optimal secara mandiri. Selain itu, metode tersebut terbukti memiliki tingkat akurasi yang tinggi dan efisien dalam proses pengklasifikasian. Seperti penelitian yang pernah dilakukan oleh Razi (2017) tentang Klasifikasi Artikel Berita Berbahasa Indonesia Menggunakan *Convolutional Neural Network* dengan nilai akurasi sebesar 96,70% dan presisi, recall serta f-measure mencapai 96,60%. Ketiga metode tersebut juga pernah dibandingkan dengan hasil dari CNN, RNN, dan HAN secara berturut turut yaitu 96,12%, 95,76%, 95,6% menggunakan dataset pertama, 87,88%, 82,63%, 87,91% menggunakan dataset kedua, dan dengan dataset ketiga menghasilkan 91,6%, 89,54%, 92,16% (Maheshwari, 2018).

Oleh karena itu, penelitian ini akan mengkomparasikan metode *convolutional neural network, recurrent neural network* dan *hierarchical attention network* untuk klasifikasi berita *hoax* dan pengelompokkan berdasarkan kategori berita*.* Penelitian ini akan menggunakan data yang bersumber dari turnbackhoax.id yang telah memvalidasi berita *hoax* yang beredar di media sosial. Data tersebut akan menjadi data latih dan juga data uji. Penelitian ini diharapkan dapat mengetahui metode *deep learning* mana yang memiliki tingkat performa yang lebih baik. Selain itu, penelitian ini juga diharapkan dapat menghasilkan sebuah sistem cerdas yang dapat mengklasifikasikan berita yang *hoax* dan yang tidak *hoax* serta dapat dikelompokkan berdasarkan kategori berita sehingga dapat diketahui kategori berita mana yang tingkat intensitas *hoax*nya tinggi dan mana yang tidak.

* 1. **Perumusan Masalah**

Berdasarkan latar belakang yang telah dipaparkan, maka dapat diambil rumusan masalah sebagai berikut:

1. Bagaimana tingkat akurasi algoritma *convolutional neural network, recurrent neural network* dan *hierarchical attention network* dalam klasifikasi berita *hoax* dan klasifikasi kategori berita?
2. Bagaimana tingkat efisiensi algoritma *convolutional neural network, recurrent neural network* dan *hierarchical attention network* dalam klasifikasi berita *hoax* dan klasifikasi kategori berita?
   1. **Batasan Masalah**

Pada penelitian ini, permasalahan dibatasi sebagai berikut:

1. Data penelitian didapatkan dari situs web turnbachoax.id yang telah memvalidasi berita *hoax* yang beredar di media sosial.
2. Data penelitian yang digunakan dari tahun 2015 sampai dengan 2019.
3. Pada penelitian ini akan dilakukan klasifikasi berita *hoax* dan tidak *hoax* serta klasifikasi kategori berita.
4. Data penelitian yang digunakan berbahasa Indonesia.
   1. **Tujuan Penelitian**

Tujuan pada penelitian ini adalah sebagai berikut:

1. Dapat mengetahui tingkat akurasi algoritma *convolutional neural network, recurrent neural network* dan *hierarchical attention network* untuk klasifikasi berita *hoax* dan ketegori berita.
2. Dapat mengetahui tingkat efisiensi algoritma *convolutional neural network, recurrent neural network* dan *hierarchical attention network* untuk klasifikasi berita *hoax* dan ketegori berita.
   1. **Manfaat Penelitian**

Dengan dilakukannya penelitian ini dapat mengetahui algoritma *deep learning* mana yang memiliki tingkat performa yang lebih tinggidalam menyelesaikan permasalahan tentang klasifikasi berita *hoax* dan kategori berita. Selain itu, diharapkan dengan adanya sistem ini dapat memudahkan masyarakat khususnya pengguna media sosial dalam memilah dan menyaring berita yang tersebar. Sehingga dampak dari penyebaran berita *hoax* dapat diminimalisir.

* 1. **Metodologi Penelitian**

Alur metodologi penelitian yang dilakukan sebagai berikut:

1. Studi Literatur

Studi literatur dilakukan dengan cara mencari dan menghimpun data ataupun sumber-sumber pustaka yang dapat mendukung penelitian serta memberikan informasi untuk menyelesaikan permasalahan pada penelitian ini. Studi kepustakaan yang digunakan bersumber dari buku, jurnal, artikel dan paper yang berkaitan pada penelitian ini.

1. Pengumpulan Data

Data yang akan digunakan bersumber dari situs turnbackhoax.id dan data tersebut sudah divalidasi.

1. Analisis Sistem

Analisis sistem ini merupakan penganalisaan terhadap kebutuhan dalam pembuatan sistem. Sehingga dapat diketahui kebutuhan-kebutuha apa saja yang diperlukan dalam pembuatan sistem pada penelitian ini.

1. Perancangan Sistem

Pada bagian perancangan ini akan melakukan pemodelan terhadap sistem yang akan dibuat berdasarkan hasil analisis yang telah dilakukan sebelumnya.

1. Implementasi Perangkat Lunak

Tahap ini akan mengimplementasikan sistem yang sudah dirancang sebelumnya.

1. Pengujian dan Analisis

Pada tahap ini akan dilakukan pengujian terhadap implementasi yang telah dikerjakan. Kemudian akan dilakukan analisis berdasarkan hasil dari pengujian.

1. Kesimpulan dan Saran

Pada penelitian ini akan diberikan kesimpulan dari hasil penelitian yang telah dilakukan dan akan menyertakan saran yang dapat dikembangkan untuk penelitian di masa depan sehingga mendapatkan hasil yang lebih baik.

* 1. **Sistematika Penulisan**

Sistematika penulisan yang digunakan dalam menyusun laporan penelitian ini adalah sebagai berikut :

Bab I Pendahuluan

Pada bagian ini membahas tentang latar belakang masalah, perumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, metodologi penelitian, dan sistematika penulisan.

Bab II Tinjauan Pustaka

Tinjauan pustaka memuat tentang dasar teori yang digunakan untuk analisis dan perancangan sistem serta implementasi pada penelitian ini. Selain itu juga sebagai bahan referensi dan pondasi untuk memperkuat argumentasi dalam penelitian ini. Teori-teori yang sesuai dengan penelitian ini antara lain media sosial, *hoax,* klasifikasi, *deep learning,* *convolutional neural network, recurrent neural network,* dan *hierarchical attention network* .

Bab III Analisis dan Perancangan Sistem

Pada bagian ini akan membahas mengenai analisa dan perancangan sistem hingga implementasi sistem dalam mengklasifikasi berita *hoax* di media sosial dan mengklasifikasi kategori berita.

Bab IV Hasil, Pengujian dan Pembahasan

Pada bab ini akan menyajikan hasil penelitian yang berisi hasil implementasi dari perancangan yang telah dibuat pada bab sebelumnya. Selain itu berisi pengujian terhadap hasil penelitian beserta pembahasannya.

Bab V Kesimpulan dan Saran

Pada bagian ini berisi kesimpulan dari hasil penelitian dan saran yang diajukan oleh penulis untuk pengembangan penelitian selanjutnya.

**Daftar Pustaka**

Afriza, A., & Adisantoso, J. (2018). Metode Klasifikasi Rocchio untuk Analisis Hoax. *Jurnal Ilmu Komputer Dan Agri-Informatika*, *5*(1), 1–10. https://doi.org/10.29244/jika.5.1.1-10

Fajri, D. (2019). Menkominfo: Ada 900 Ribu Situs Penyebar Informasi Hoax. Retrieved June 29, 2019, from www.okezone.com website: https://news.okezone.com/read/2019/02/14/337/2018062/menkominfo-ada-900-ribu-situs-penyebar-informasi-hoax

Lhaksmana, K. M., Nhita, F., & Budhiarto, A. (2017). Klasifikasi Pengguna Media Sosial Twitter Dalam Persebaran Hoax Menggunakan Metode Backpropagation Classification of Users Social Media Twitter in the Hoax Spread. *E-Proceeding of Engineering*, *4*(2), 3082–3090.

Librianty, A. (2017). Survei: Media Sosial Jadi Sumber Utama Penyebaran Hoax. Retrieved September 11, 2019, from www.liputan6.com website: https://www.liputan6.com/tekno/read/2854713/survei-media-sosial-jadi-sumber-utama-penyebaran-hoax

Maheshwari, A. (2018). *Report on Text Classifiction using CNN, RNN & HAN*. Retrieved from https://medium.com/jatana/report-on-text-classification-using-cnn-rnn-han-f0e887214d5f

Maulina, D., & Sagara, R. (2018). Klasifikasi Artikel Hoax Menggunakan Support Vector Machine Linear Dengan Pembobotan Term Frequency – Inverse Document Frequency. *Jurnal Mantik Penusa*, *2*(1), 35–40.

Tanjung, B. S. (2018). Pendekatan Text Mining sebagai Sistem Pendeteksi Pemberitaan Palsu yang Tersebar dalam Twitter. *Universitas Widya Kartika*, 1–6.

Wearesocial. (2019). Digital 2019: Indonesia. Retrieved June 29, 2019, from www.datareportal.com website: https://datareportal.com/reports/digital-2019-indonesia/

Zulfa, I., & Winarko, E. (2017). Sentimen Analisis Tweet Berbahasa Indonesia Dengan Deep Belief Network. *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, *11*(2), 187–198. https://doi.org/10.22146/ijccs.24716