

BAB I

PENDAHULUAN

Bab ini, menjelaskan latar belakang, identifikasi masalah, rumusan masalah, ruang lingkup kegiatan, tujuan, dan sistematika penulisan.

1.1 Latar Belakang

Dalam Kamus Besar Bahasa Indonesia (KBBI) kata *hoax* diserap menjadi hoaks yang diartikan sebagai informasi bohong. Kata *hoax* sendiri baru mulai digunakan sekitar tahun 1808 (Aditiawarman, 2019). Kata tersebut dipercaya berasal dari kata *hocus* yang berarti untuk mengelabui. Kata *hocus* sendiri merupakan singkatan dari *hocus pocus*, sejenis mantra yang kerap digunakan dalam pertunjukan sulap. Namun berita *hoax* adalah sebuah usaha menipu pembaca atau pendengar untuk mempercayai sesuatu padahal pembuat berita tersebut tahu bahwa informasi tersebut adalah palsu atau bohong (Rahadi, 2017). Penyebaran informasi *hoax* bertujuan sebagai bahan lelucon, iseng, dan biasanya untuk menjatuhkan pesaing (*black campaign*). Dampak yang dihasilkan oleh *hoax* merupakan dampak yang tidak bisa disadari secara langsung, karena dapat menyerang pemikiran pembacanya dan jika tidak berhati-hati dapat mempengaruhi cara berpikir pembacanya. Menurut Alessandro Bondielli istilah *hoax* biasanya disebut sebagai “virus pikiran”, hal ini dikarenakan kemampuannya untuk mereplikasi diri, mengadaptasi, memutasi dan bertahan di dalam pikiran manusia (Alessandro Bondielli a, 2019).

Saat ini media sosial merupakan media komunikasi yang efektif, dan efisien. Penggunaan media sosial sebagai jembatan untuk membantu proses peralihan masyarakat yang tradisional ke masyarakat yang modern, khususnya untuk mentransfer informasi pembangunan yang dilaksanakan pemerintah kepada masyarakatnya. Sebaliknya masyarakat dapat menyampaikan informasi langsung kepada pemerintah tentang berbagai hal terkait dengan pelayanan yang diterima. Media sosial telah menjadi alat penerbitan penting bagi jurnalis dan metode konsumsi utama bagi masyarakat yang mencari berita terbaru (Arkaitz Zubiaga, 2018). Jurnalis dapat menggunakan media sosial untuk melaporkan pendapat publik tentang berita baru, dan bahkan untuk menemukan berita baru yang potensial, sedangkan masyarakat dapat mengikuti perkembangan berita baru. Namun dengan

tidak adanya kontrol dan pemeriksaan fakta atas pos-pos media sosial menjadikan media ini subur untuk penyebaran informasi yang tidak diverifikasi. Orang-orang sering mempublikasikan atau membagikan postingan orang lain. Seringkali, tajuk yang menarik cukup untuk sebuah artikel untuk dibagikan ribuan kali, meskipun konten yang mungkin tidak berdasar atau palsu. Pengaruh buruk dari *hoax* dapat mempengaruhi seseorang melalui sebuah tulisan sehingga dengan *hoax* tersebut dapat mempengaruhi pikiran waras seseorang, sedangkan melalui gambar dapat memunculkan rasa takut dan terancam. Menurut hasil riset Lembaga Ilmu Pengetahuan Indonesia (LIPI), masyarakat yang fanatik lebih mudah terkena *hoax*. Jika dibiarkan berita *hoax* bisa sangat persuasif, maka diperlukan mengembangkan strategi untuk mengidentifikasi berita *hoax* yang disebarkan di media sosial (Anshar, 2019).

Oleh karena itu penting untuk mengembangkan sebuah teknik yang mampu mengidentifikasi berita *hoax* berbahasa Inggris pada media sosial. Dengan memanfaatkan metode *Deep Learning* khususnya *Natural Language Processing (NLP)*. *Deep learning* sendiri mempunyai arsitektur antara lain *Deep Feedforward Network*, *Recurrent Neural Network*, *Convolutional Neural Network* yang merupakan pengembangan dari jaringan syaraf tiruan untuk memberikan ketepatan tugas seperti deteksi objek, pengenalan suara, terjemahan bahasa dan lain – lain (Geoffrey E. Hinton, 2006).

Deep Learning merevolusi bidang-bidang seperti pemrosesan bahasa alami atau *Natural Language Processing* seperti penelitian Adhitya Rao memakai metode *Recurrent Neural Network* variasi LSTM untuk klasifikasi data sosial media tentang kecenderungan mendukung partai politik Demokrat atau partai politik Republik di Amerika Serikat, LSTM menunjukkan tingkat akurasi sebesar 87.57% (Adithya Rao, 2016).

Berita *hoax* mengacu pada cerita palsu yang disengaja disebarluaskan melalui internet dan media sosial. Berita seperti ini bisa jadi sangat persuasif, maka dari itu dibutuhkan strategi untuk mengidentifikasi secara kritis berita yang dibaca pada media sosial. Pada penelitian O.E. Taylor melakukan identifikasi berita *hoax* pada media sosial. Pada dataset *fake_or_real_news* di proses menggunakan algoritma *machine learning*. Hasil performa yang ditunjukkan (O. E. Taylor, 2020)

dengan menggunakan metode *Logistic Regression* adalah akurasi sebesar 92%, kemudian menggunakan metode *MultinomialNaiveBayes* menghasilkan akurasi sebesar 89%, dan yang terakhir menggunakan metode *Support Vector Classifier* menghasilkan akurasi sebesar 87%.

Oleh karena itu pada penelitian ini merancang sebuah sistem untuk mengidentifikasi berita *hoax* berbahasa Inggris pada media sosial dengan menggunakan metode *RNN* memakai variasi *LSTM* untuk menghitung tingkat keberhasilan akurasi nya.

1.2 Rumusan Masalah

Berdasarkan latar belakang masalah yang dijelaskan diperoleh rumusan masalah sebagai berikut :

1. Bagaimana merancang suatu sistem yang dapat mendeteksi berita *hoax* berbahasa Inggris pada media sosial?
2. Bagaimana metode *LSTM* diimplementasikan untuk membangun model prediksi berita *hoax* berdasarkan dataset yang diberikan ?

1.3 Tujuan

Tujuan penelitian ini adalah membangun model dengan mengimplementasikan metode *Long Short-Term Memory* untuk mengidentifikasi berita *hoax*.

1.4 Ruang Lingkup

Dalam penelitian ini dilakukan, dibuat ruang lingkup penelitian agar cakupan penelitian lebih spesifik. Adapun penelitian yang dibuat sebagai berikut :

1. Sistem berfokus mengidentifikasi berita *hoax* jenis *fake news*.
2. Sistem yang dibangun mengidentifikasi berita *hoax* berbahasa Inggris.
3. Sistem yang dibangun menggunakan *Glove 50d*.
4. Dataset yang digunakan diperoleh dari website *kaggle.com*.
5. Dataset pengujian diambil dari dataset *kaggle.com* "*fake_or_real_news*" yang telah dipisahkan sebelum dilakukan pelatihan model.

1.5 Metode Penelitian

Metode yang digunakan pada penelitian ini merupakan metode dengan menerapkan algoritma *deep learning text classification LSTM* untuk mengidentifikasi berita *hoax* berbahasa Inggris pada media sosial berdasarkan teks berita yang teridentifikasi *hoax* yang didukung oleh pendekatan pengembangan *software* model *prototype*. Tahapan penelitian yang dilakukan mengikuti alur model *prototype* yang dimulai dari identifikasi kebutuhan baik *software* maupun *hardware*, desain, pembuatan *prototype*, evaluasi, dan penyempurnaan *prototype*.

Dalam melakukan klasifikasi pada teks, hal pertama yang dilakukan adalah *training* terhadap dataset berita *hoax* dan berita *real (non hoax)* berbahasa Inggris yang telah diberi label berdasarkan model *LSTM* untuk dihasilkan model yang menyimpan bobot dari dataset.

1.6 Tinjauan Pustaka

Dalam penelitian ini melibatkan beberapa pustaka yang berkaitan dengan penelitian yang dilakukan adalah :

(Costel-Sergiu Atodiresei*, 2018) penelitian ini menjelaskan tentang identifikasi berita palsu pengguna palsu pada salah satu media sosial yaitu twitter dengan menggunakan NLP Tools (NER – OpenCalais, Sentiment140) dengan hasil telah berhasil mendeteksi berita palsu dan pengguna palsu.

(Gang Liu, 2018) pada penelitian ini menjelaskan tentang klasifikasi teks, ekstraksi fitur dan desain classifier sangat penting. *LSTM* telah menunjukkan kinerja yang lebih baik di banyak dunia nyata. Namun, masih sulit untuk memahami semantic, keakuratan dan klasifikasi masih perlu ditingkatkan. Untuk mengatasi masalah ini, makalah ini menyajikan metode *LSTM* yang ditingkatkan, yaitu *AC-BiLSTM*, di mana lapisan konvolusional, *BiLSTM* dan mekanisme perhatian digunakan untuk meningkatkan pemahaman semantik dan meningkatkan klasifikasi 790 akurasi.

(Maaz Amajd, 2017) penelitian ini menjelaskan text classification dengan Deep Neural Network yaitu *LSTM* memberikan tingkat akurasi tertinggi yaitu 88.03% dalam pengklasifikasian yang bersifat sekuensial pada dataset IMDB

dibandingkan dengan CNN yang memiliki tingkat akurasi 87.71% dan Naive-Bayes dengan tingkat akurasi 83.29%.

(Ji Young Lee, 2016) penelitian ini menjelaskan dalam artikel ini menyajikan pendekatan berbasis JST Recurrent dan Convolutional Neural Network untuk klasifikasi teks pendek berurutan. Penelitian ini menunjukkan bahwa menambahkan informasi sekuensial meningkatkan kualitas prediksi, dan kinerja tergantung pada apa informasi yang digunakan dalam model. Model penelitian ini mencapai yang terbaik hasil pada tiga dataset berbeda yaitu DSTC 4 66.2 MRDA 84.6 SwDA 73.1.

(Bajaj, 2017) penelitian ini menjelaskan bahwa arsitektur *RNNs* menghasilkan hasil yang sangat menarik. Tercatat bahwa beberapa artikel berita dalam dataset panjangnya ribuan kata, dan bahkan *RNN* yang dilengkapi dengan *LSTM* atau *GRU*, tidak dapat menyimpan memori lebih dari seratus kali tangga. Rahasia kesuksesan *RNNs* dalam tugas ini mungkin ada hubungannya dengan fakta bahwa jaringan menggunakan kondisi internal yang besar dengan ukuran 100, sering kali menimbulkan tantangan dalam pelatihan dan penggunaan sumber daya.

(Adithya Rao, 2016) penelitian ini menjelaskan penerapan Word embeddings dan *LSTM* untuk text classification yang dimana diputuskan oleh konteks aplikasi. Untuk aplikasi pertama dalam konteks dukungan pelanggan di media social dimana agen perusahaan menanggapi keluhan pelanggan. Dengan menyelesaikan setiap Bahasa secara terpisah dan menemukan modelnya dengan hasil 90% untuk akurasi sebagian besar Bahasa. Untuk aplikasi kedua dalam konteks mengklasifikasi pesan sebagai demokasi atau republic dengan hasil akurasi 87%.

(Aini Suri Talita¹, 2019) penelitian ini menjelaskan penerapan algoritma Long Short-Term Memory *LSTM* untuk mendeteksi ujaran kebencian (Hate Speech). Tahapan dalam penelitian ini dimulai dari studi kepustakaan, pengumpulan data, pra-pemrosesan data, pembuatan model word2vec, perancangan model pendeteksi ujaran kebencian, pelatihan model, dan pengujian model. Model ini dilatih dengan menggunakan 950 kalimat dan diuji dengan 190 kalimat dari dataset yang bersumber dari media sosial Facebook. Hasil yang didapatkan dari ini

dengan menggunakan data testing 190 kalimat dari 950 kalimat dari dataset, algoritma Long Short Term Memory sudah cukup baik dalam mendeteksi kalimat ujaran kebencian dengan nilai parameter recall mencapai 0.7021. Untuk nilai parameter lainnya masih cenderung rendah yang mencakup bahasa informal, dan diambil langsung dari media sosial dimana penulisan dari kata-kata seringkali berubah karena disingkat, banyaknya ejaan yang tidak benar atau tidak konsisten, juga kecenderungan mengganti huruf tertentu sebagai angka sehingga tidak dapat terdeteksi sebagai ujaran kebencian sesuai dengan kriteria pada model yang telah dibangun.

(Pengfei Liu, 2016) penelitian ini menjelaskan text klasifikasi menggunakan Recurrent Neural Network dengan tiga arsitektur yang berbeda yaitu Uniform-Layer Architecture, Coupled-Layer Architecture, dan Shared-Layer Architecture. Hasil yang didapatkan adalah meningkatnya kinerja sekelompok tugas terkait dengan menjelajahi fitur-fitur umum.

(Oluwaseun Ajao, 2018) penelitian ini menjelaskan tentang klasifikasi berita palsu pada posting twitter menggunakan hybrid convolutional neural network dan long short-term recurrent neural network. Hasil yang didapatkan dalam klasifikasi dalam mendeteksi berita palsu sebesar 82% akurasi dengan model *LSTM* vanilla dalam hal ketepatan, daya ingat dan pengukuran FM. Untuk *CNN* dan *RNN* membutuhkan kumpulan data yang lebih besar.

(Georgios Gravanis, 2019) penelitian ini melakukan perbandingan deteksi berita palsu dimana fitur yang diusulkan dikombinasikan dengan algoritma machine learning memperoleh akurasi hingga 95% dari semua dataset yang digunakan dengan AdaBoost menjadi yang pertama pada peringkat dan algoritma *SVM* & *Bagging* berada di peringkat berikutnya. Hasil tersebut membuktikan bahwa klasifikasi artikel berdasarkan kebenarannya memungkinkan memilih fitur yang tepat pada algoritma machine learning yang sesuai.

1.7 Sistematika Penulisan

Berikut adalah sistematika penulisan yang digunakan untuk memberikan gambaran isi dari laporan ini:

- BAB I** **PENDAHULUAN**
- Pada bab ini akan dijelaskan mengenai latar belakang masalah, rumusan masalah, ruang lingkup, tujuan, tinjauan pustaka, kontribusi penelitian dan sistematika penulisan.
- BAB II** **LANDASAN TEORI**
- Pada Bab ini akan dijelaskan teori-teori dasar yang digunakan pada penelitian implementasi metode *Long Short-Term Memory* untuk identifikasi berita *hoax* berbahasa inggris pada media social.
- BAB III** **METODE PENELITIAN**
- Pada bab ini akan dijelaskan metode yang digunakan dalam peneitian. Di tahapan ini terdapat blok diagram, *work flow*, *flowchart*, dan sejenis lainnya untuk menjelaskanpercangan yang dilakukan.
- BAB IV** **IMPEMENTASI PENGUJIAN**
- Pada Bab ini ditampilkan hasil perancangan yang diajukan. Akan ditampilkan hasil penelitian implementasi metode *Long Short-Term Memory* untuk identifikasi berita *hoax* berbahasa inggris pada media social yang diajukan dan pembangunan aplikasi.
- BAB V** **PENUTUP**
- Pada bab ini terdapat kesimpulan dan saran dari hasil penelitian.