

BAB I

PENDAHULUAN

1.1 Latar Belakang

Berita palsu adalah informasi yang direkayasa dan dibuat dengan tujuan untuk menutupi informasi asli dengan menggunakan bahasa yang meyakinkan tapi tidak dapat diverifikasi kebenarannya (Gumilar, 2017). Berita palsu biasanya dibuat dengan menggunakan judul yang provokatif dan mengandung fitnah. Beredarnya berita palsu ini, biasanya didasari oleh dua hal, yaitu ekonomi dan politik (Nugroho, 2017). Motif ekonomi dalam penyebaran berita palsu biasanya digunakan untuk membuat berita yang sensasional agar dapat menarik banyak pengunjung untuk mengakses situs berita tersebut sehingga dapat menghasilkan banyak keuntungan. Selain itu, berita palsu juga biasanya dibuat untuk menarik massa untuk menjatuhkan lawan politik dengan menyebarkan fitnah terhadap lawan politik tersebut. Contoh berita palsu yang beredar dapat dilihat pada Gambar 1.1.

[SALAH] Anies Akui Semua Hutangnya, Permalukan Diri & Para Pendukungnya

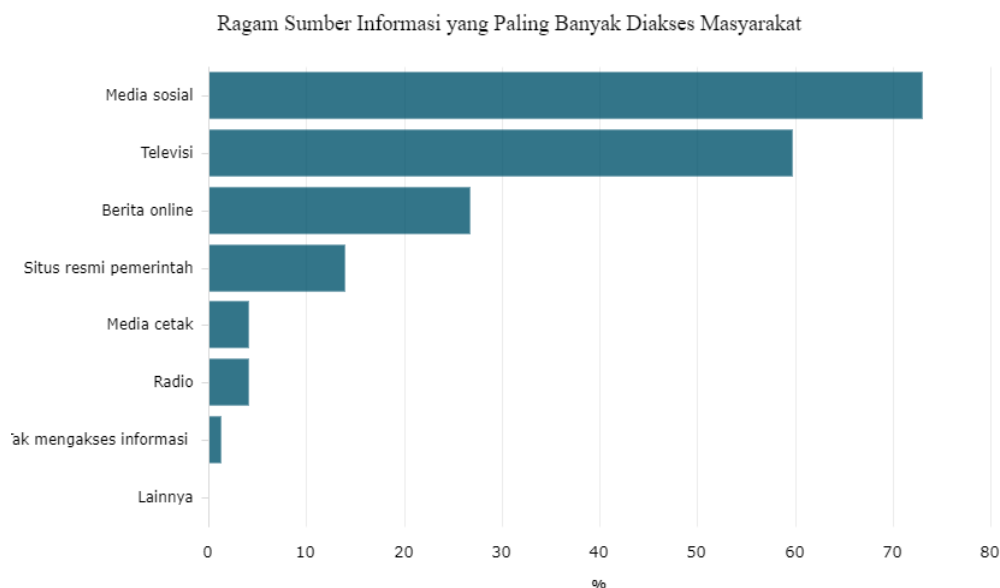


Gambar 1.1 Contoh berita palsu (kominfo.go.id)

Dampak yang dihasilkan oleh berita hoaks seringkali tidak disadari oleh pembaca, karena berita hoaks dapat menyerang pola pikir pembaca sehingga pembacanya sangat mempercayai berita tersebut. Istilah hoaks terkadang disebut

juga sebagai “*virus of the mind*”, karena kemampuannya yang dapat beradaptasi dan bermutasi dalam pikiran manusia (Vuković dkk., 2009). Jika terus dibiarkan, penyebaran berita hoaks dapat berdampak pada bidang sosial, ekonomi, keamanan hingga mengancam keutuhan negara.

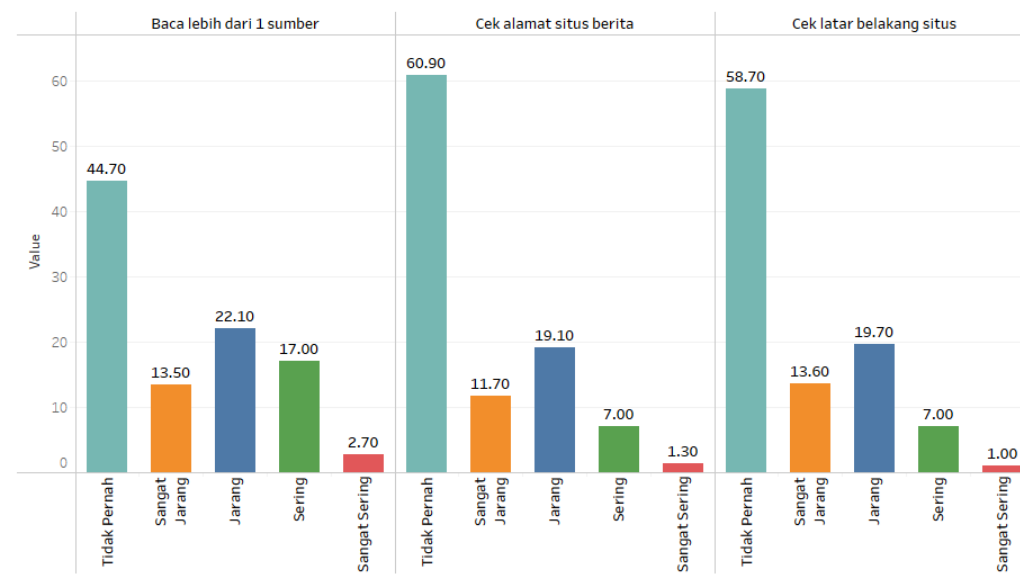
Penyebaran berita palsu dapat dilakukan oleh perorangan maupun kelompok. Menurut laporan Kementerian Komunikasi dan Informatika (KEMKOMINFO), terdapat 565.449 konten hoaks yang tersebar di media sosial sepanjang tahun 2021. Banyaknya berita palsu yang beredar tidak lepas dari pengaruh perkembangan teknologi dan media sosial yang membuat masyarakat dapat mengakses berita kapanpun dan dimanapun. Katadata Insight Center (KIC) melakukan survei terkait sumber informasi yang paling sering diakses oleh masyarakat Indonesia (Annur, 2022). Hasil survei tersebut dapat dilihat pada Gambar 1.2.



Gambar 1.2 Ragam sumber informasi yang paling sering diakses masyarakat (KIC)

Gambar 1.2 menampilkan hasil survei yang dilakukan oleh Katadata *Insight Center* (KIC) terkait sumber informasi yang paling sering diakses oleh masyarakat. Dari hasil survei tersebut dapat diketahui bahwa sebanyak 73% dari 10.000 responden mendapatkan informasi dari media sosial, dan sebanyak 26.7% dari portal berita online. Hal ini membuktikan bahwa perkembangan teknologi dan jaringan internet sangat berperan penting dalam penyebaran informasi. Namun, banyaknya informasi yang disebarkan tersebut, perlu dikonfirmasi kembali terkait kebenarannya untuk menghindari terjadinya penyebaran berita palsu. Menurut (Nugroho, 2017)

terdapat lima langkah yang dapat dilakukan untuk mencegah penyebaran berita hoaks, yaitu: Waspada terhadap judul yang provokatif, cek situs berita, periksa fakta dengan membaca lebih dari satu sumber, cek keaslian foto yang digunakan dalam artikel berita, dan ikut serta dalam komunitas anti hoaks. Terkait langkah pencegahan penyebaran berita hoaks tersebut, Kementerian Komunikasi dan Informatika (KEMKOMINFO) juga melakukan survei untuk melihat tingkat kewaspadaan masyarakat dalam menerima berita secara online (Ahdiat, 2022). Survei tersebut dilakukan pada 10.000 responden yang berusia 13-70 tahun. Hasil survei dapat dilihat pada Gambar 1.3.



Gambar 1.3 Kebiasaan masyarakat dalam mengonsumsi berita online (KIC)

Berdasarkan hasil survei pada Gambar 1.3, diketahui bahwa mayoritas masyarakat Indonesia tidak pernah melakukan pemeriksaan terkait sumber berita yang diperoleh secara online dan cenderung hanya membaca berita berdasarkan satu sumber saja tanpa mengecek kebenaran berita dari sumber lain. Hal ini dapat menyebabkan masyarakat rentan menyebarkan berita palsu secara tidak sadar. Oleh karena itu, dibutuhkan sebuah sistem yang dapat mendeteksi berita palsu berdasarkan narasi yang terdapat pada berita yang beredar, untuk membantu masyarakat dalam mendeteksi konten berita yang terindikasi sebagai berita palsu atau hoaks.

Penelitian terkait sistem pendeteksi berita palsu telah banyak dilakukan oleh peneliti dari berbagai belahan dunia. Diantaranya yaitu, penelitian yang dilakukan oleh (Jadhav & Thepade, 2019), penelitian ini menggunakan algoritma *Recurrent Neural Network* (RNN) yang ditingkatkan dengan *Deep Structured Semantic Model* (DSSM) untuk membuat sistem pendeteksi berita palsu. Data yang digunakan dalam penelitian ini adalah dataset LIAR yang berjumlah 10239 data dan dapat diakses secara bebas untuk digunakan sebagai data penelitian. Dengan menggunakan framework tersebut, penelitian ini berhasil membuat sebuah sistem pendeteksi berita palsu dengan tingkat akurasi sebesar 99%. Penelitian lain terkait sistem deteksi berita palsu juga pernah dilakukan oleh (Granik & Mesyura, 2017), penelitian ini membuat sebuah sistem pendeteksi berita palsu dengan menggunakan algoritma naive bayes classifier dan menerapkannya pada data post berita di media sosial facebook. Penelitian ini menggunakan facebook API untuk mengambil data dari facebook dan menggunakan pendekatan artificial intelligence (AI) untuk mendeteksi berita palsu yang beredar. Hasilnya, sistem ini dapat melakukan klasifikasi berita palsu dengan tingkat akurasi mencapai 74%. Penelitian lain juga pernah dilakukan oleh (Kaliyar dkk., 2020) yang menggunakan pendekatan Deep Convolutional Neural Network (FNDNet) untuk membuat sistem pendeteksi berita palsu. Data yang digunakan dalam penelitian ini adalah dataset berita palsu yang beredar selama masa pemilihan presiden US pada tahun 2016. Penelitian ini berhasil membuat sebuah sistem pendeteksi berita palsu dengan tingkat akurasi sebesar 98.36%.

Long Short-Term Memory (LSTM) adalah pengembangan dari algoritma *Recurrent Neural Network* (RNN) yang dapat menyimpan informasi dalam jangka waktu yang panjang. Algoritma yang dikembangkan oleh Hochreiter dan Schmidhuber pada tahun 1997 ini pertama kali diperkenalkan untuk mengatasi permasalahan *vanishing gradient problems* pada RNN (Raschka & Mirjalili, 2019). Algoritma RNN memiliki kinerja yang sangat baik dalam melakukan klasifikasi pada berita hoaks (Jadhav & Thepade, 2019). Sementara itu, algoritma LSTM dikembangkan berdasarkan algoritma RNN dengan mempertahankan setiap kelebihan, dan mengatasi kelemahannya sehingga algoritma LSTM dapat menghasilkan hasil prediksi yang lebih akurat. Penelitian ini menggunakan metode LSTM dengan *model pretrained Indobert* untuk membuat sebuah sistem prediksi

berita palsu pada konten berita berbahasa Indonesia. Model *pretrained Indobert* digunakan untuk melakukan *word embedding* pada data *input*. Model *pretrained* digunakan karena dapat meningkatkan kinerja model untuk tugas klasifikasi teks dengan mentransfer hasil pelatihan model sebelumnya pada model yang dibuat (J. Liu & He, 2020). Penelitian ini bertujuan untuk membuat sebuah sistem pendeteksi berita palsu dan menguji kinerja dari sistem tersebut.

Berdasarkan penjelasan mengenai penggunaan LSTM dan model *pretrained Indobert* dalam pembuatan sistem pendeteksi berita palsu menggunakan LSTM dan model *pretrained Indobert*, dirumuskan hipotesis h_0 dan h_1 sebagai berikut:

h_0 : Terdapat perbedaan yang signifikan dalam kinerja sistem pendeteksi berita palsu antara penggunaan LSTM dan RNN dengan penggunaan *model pretrained Indobert*.

h_1 : Tidak ada perbedaan yang signifikan dalam kinerja sistem pendeteksi berita palsu antara penggunaan LSTM dan RNN dengan penggunaan *model pretrained Indobert*.

Untuk mendukung kedua hipotesis diatas, dibuatlah premis sebagai berikut:

1. LSTM sebagai Model Dasar:
 - a. LSTM adalah salah satu jenis arsitektur jaringan saraf yang dapat digunakan untuk tugas pemrosesan teks seperti klasifikasi berita palsu.
 - b. LSTM merupakan pengembangan dari Algoritma RNN yang dibuat untuk mengatasi permasalahan *vanishing gradient*.
2. *Model Pretrained Indobert* untuk *word embedding*:
 - a. *Model pretrained Indobert* adalah model bahasa yang telah dilatih pada data bahasa Indonesia yang besar dan dapat melakukan pemahaman bahasa dengan baik.
 - b. *Model pretrained Indobert* memiliki kemampuan untuk menangani konteks bahasa Indonesia yang kompleks
3. Kinerja Model:
 - a. Untuk menguji hipotesis, akan dilakukan eksperimen dengan menggunakan dataset yang telah dikumpulkan.
 - b. Kinerja sistem pendeteksi berita palsu akan diukur menggunakan PR Curve untuk menghitung nilai *precision*, *recall*, dan *f-1 score*.
 - c. Berdasarkan hasil pengujian sistem, akan dibuat perbandingan kinerja LSTM dan *model pretrained Indobert* dalam mendeteksi berita palsu.

1.2 Rumusan Masalah

Berdasarkan latar belakang masalah yang telah dijelaskan sebelumnya, penelitian ini memiliki beberapa rumusan masalah, yaitu sebagai berikut:

1. Bagaimana merancang model untuk sistem pendeteksi berita palsu berbahasa Indonesia menggunakan metode LSTM?
2. Bagaimana menerapkan metode LSTM untuk mendeteksi berita palsu berbahasa Indonesia?
3. Bagaimana kinerja metode LSTM untuk mendeteksi berita palsu berbahasa Indonesia?

1.3 Tujuan Penelitian

Berdasarkan rumusan masalah yang telah dijelaskan di atas, penelitian ini bertujuan untuk:

1. Merancang model sistem pendeteksi berita palsu berbahasa Indonesia menggunakan metode LSTM;
2. Mengimplementasikan metode LSTM untuk membuat sistem pendeteksi berita palsu;
3. Mengevaluasi dan menganalisis kinerja metode LSTM dalam mendeteksi berita palsu berbahasa Indonesia.

1.4 Manfaat Penelitian

Penelitian ini memiliki manfaat sebagai berikut:

1. Membantu masyarakat dan pemerintah Indonesia dalam menangani penyebaran berita hoax;
2. Terkumpulnya dataset berita hoax berbahasa Indonesia yang sudah tervalidasi dan dapat digunakan untuk penelitian terkait berita hoaks;
3. Mengetahui pemodelan dan kinerja dari sistem pendeteksi berita palsu dengan menggunakan metode LSTM.

1.5 Batasan Masalah

Penelitian ini memiliki beberapa batasan masalah agar penelitian yang dilakukan menjadi lebih terarah. Batasan masalah pada penelitian ini antara lain:

1. Data yang digunakan merupakan data teks narasi berita berbahasa Indonesia
2. Data berita non hoaks yang dikumpulkan berasal dari beberapa portal berita terpercaya berdasarkan survey KIC (Annur, 2023), dan juga website resmi kementerian Indonesia, yaitu website Kementerian Kesehatan dan Kementerian Agama Indonesia.
3. Data berita hoaks yang dikumpulkan merupakan berita yang terbit pada tahun 2019 sampai tahun 2023.